East European Law Research Center

Special Issue
AI and Law
2025

PEER REVIEWED JOURNAL

ACCESS TO JUSTICE IN EASTERN EUROPE

ACCESS TO JUSTICE IN EASTERN EUROPE

Founded by the East European Law Research Center

AJEE is an English-language journal which covers issues related to access to justice and the right to a fair and impartial trial. AJEE focuses specifically on law in eastern European countries, such as Ukraine, Poland, Lithuania, and other countries of the region, sharing in the evolution of their legal traditions. While preserving the high academic standards of scholarly research, AJEE allows its contributing authors, especially young legal professionals, and practitioners, to present their articles on the most current issues.

For further information on our activities and services, please visit our website http://ajee-journal.com
To submit your manuscript, please follow the instructions in our Guide and use our Scholastica submission platform. For any queries, or if you need further assistance, feel free to contact us at info@ajee-journal.com or editor@ajee-journal.com.

ISSN 2663-0575
Publishing House 'Academic Insights Press'
https://academicinsightspress.com
info@academicinsightspress.com

Access to Justice in Eastern Europe

Volume 8
Special Issue 'AI and Law'
December 2025

TABLE OF CONTENTS

REVIEW ARTICLES

OPINION ARTICLES

CASE STUDIES

Access to Justice in Eastern Europe
ISSN 2663-0575 (Print)
ISSN 2663-0583 (Online)
http://ajee-journal.com

EDITORIAL

Editorial

FROM EDITORS: ARTIFICIAL INTELLIGENCE AND LAW IN A WORLD OF ACCELERATING CHANGE

Goda Strikaitė-Latušinskaja and Yuliia Hartman

ABSTRACT

This editorial introduction to the AJEE Special Issue on Artificial Intelligence and Law situates current technological developments within a rapidly evolving global environment where the growing use of digital and algorithmic tools intersects with fundamental rights, public governance, and the core functions of justice systems. The text highlights how artificial intelligence (AI) challenges long-standing legal assumptions (about accountability, procedural fairness, and institutional design), while simultaneously offering new opportunities for efficiency, innovation, and access to justice. The contributions to this Special Issue examine these questions across diverse jurisdictions: from Kazakhstan's emerging AI regulatory framework and the complexities of AI-assisted adjudication to civil-law challenges in Arab legal systems, criminal justice developments in Germany and Ukraine, and evolving digital rights governance in Saudi Arabia. Together, these works underscore that the legal and institutional questions raised by AI cannot be confined to any single discipline or national setting; instead, they require an interdisciplinary, comparative, and human-rights-centred approach attuned to local realities and global standards. This Introduction invites scholars and practitioners to reflect on the conditions under which AI can strengthen, rather than erode, the rule of law and public trust.

DOI:
https://doi.org/10.33327/AJEE-18-8.S-ed000164

1 FROM THE GUEST EDITOR: GODA STRIKAITĖ-LATUŠINSKAJA

Dear readers,

We stand at a moment when technological innovation is reshaping the structure and operation of legal systems with unprecedented speed. Artificial intelligence, now recognised as a key enabling technology[1], has moved from the periphery of innovation debates to the centre of legal and policy agenda. Artificial intelligence has increasingly become embedded in public administration and justice systems, reflecting broader EU efforts to digitalise public services,[2] streamline administrative and judicial processes,[3] and promote the responsible use of emerging technologies across legal institutions.[4] The question is no longer whether AI will transform law, but how deeply and under what safeguards it will influence rights, duties, and institutional legitimacy across jurisdictions.

The adoption of the EU Artificial Intelligence Act,[5] the world's first binding legal framework for AI, marks a milestone in steering these developments. Yet many of its obligations will apply only gradually. This transitional landscape reinforces the growing reliance on soft-law instruments, such as ethical guidelines, recommendations, and institutional policies that increasingly guide technological practice while binding rules continue to evolve.[6] In modern technology governance, such soft-law sources have

1 The 2018-dated report of the High Level Group on Industrial Technologies recognised AI as a 'key enabling technology' highlighting the transformative role of AI and the necessity for the industry to use AI to maintain its leadership, see European Commission, *Re-finding Industry – Defining Innovation: Report of the independent High Level Group on industrial technologies* (Publications Office of the EU 2018) 20. doi:10.2777/927953.

2 European Commission, *A Digital Single Market Strategy for Europe* COM(2015) 192 final; European Commission, *Shaping Europe's Digital Future* COM(2020) 67 final; European Commission, *EU eGovernment Action Plan 2016–2020: Accelerating the digital transformation of government* COM(2016) 179 final.

3 CEPEJ, *The Use of Information Technologies in European Courts* (2008; 2018); CEPEJ, *European Judicial Systems – Efficiency and Quality of Justice* (2006–2020 series).

4 European Parliament Resolution of 20 October 2020 with Recommendations to the Commission on a Framework of Ethical Aspects of Artificial Intelligence, Robotics and Related Technologies [2021] OJ C 404/63; European Commission, *Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts* COM(2021) 206 final.

5 Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying down Harmonised Rules on Artificial Intelligence and Amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) [2024] OJ L 2024/1689.

6 See for example, High-Level Expert Group on AI, *Ethics Guidelines for Trustworthy AI* (European Commission 2019); CEPEJ, *The European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment* (adopted 3–4 December 2018); CEPEJ, *European Judicial Systems: CEPEJ Evaluation report, 2024 Evaluation cycle (2022 data)* (European Commission 2024).

become an essential complement to legislation, shaping expectations and safeguards well before formal obligations take effect.[7]

This Special Issue of *Access to Justice in Eastern Europe* brings together research from diverse jurisdictions united by a shared concern: reconciling technological innovation with justice, accountability, and human dignity. The contributions collectively illuminate how different legal systems are responding to the opportunities and risks posed by AI in public governance, adjudication, and private law.

The article **"AI and Law: Procedural Safeguards and Regulatory Challenges in Kazakhstan"** examines the emerging regulatory framework in Kazakhstan and reveals significant gaps in the governance of high-risk systems. Drawing on EU and OECD models, the authors propose a phased reform strategy: targeted legislative amendments in the short term, followed by a dedicated AI law to strengthen transparency, oversight, and accountability.

The article **"Artificial Intelligence in Courts and Dispute Resolution: Challenges and Opportunities"** turns to the sensitive domain of adjudication. Through doctrinal and comparative analysis, it distinguishes between assistive AI tools and fully autonomous decision-making, concluding that only the former can be reconciled with fair trial guarantees. The authors highlight the need for human oversight, transparency, and safeguards against bias to preserve judicial independence and public trust.

The article **"Legal Challenges Related to Contractual Negotiations via AI Technologies"** explores the civil-law implications of AI systems used in contractual negotiations within Arab jurisdictions. Identifying a regulatory vacuum in Egypt and the UAE, it contrasts these frameworks with emerging EU liability models and proposes clearer definitions, evidentiary rules, and supervisory mechanisms to address complex issues of fault and causation.

Together, these articles demonstrate how AI is testing and, in some cases, reshaping legal doctrine and institutional design. Their insights underscore the need for coherent, rights-centred regulatory approaches that can accommodate technological complexity while safeguarding accountability and public trust.

7 European Commission, *Building Trust in Human-Centric Artificial Intelligence* COM/2019/168 final.

2 FROM MANAGING EDITOR: Yuliia Hartman

Dear readers,

Interdisciplinarity has become a defining trend in contemporary research. It marks a shift in traditional disciplines as they adapt to remain relevant and deliver a tangible impact on today's complex societal challenges. By integrating knowledge across fields, interdisciplinary research enables scholars to revisit rigid or traditional concepts from new perspectives, often leading to groundbreaking insights and transformative developments.

Law, as one of the oldest pillars of the classical sciences, is often perceived as rigid, formalistic, and resistant to change. Its methods and language seem firmly standardised, and its application appears constrained within established doctrinal boundaries. This Special Issue aims to challenge that stereotype by demonstrating how legal scholarship increasingly intersects with other rapidly developing scientific domains.

Today's realities offer numerous opportunities to combine legal studies with other disciplines. One of the most promising intersections is the relationship between law and artificial intelligence, a field that draws substantial academic interest from researchers in mathematics and information technologies. The symbiosis between law and AI offers a new paradigm for legal research, particularly in processing and analysing massive datasets accumulated over long periods.

The intersection of law and technology is no longer a theoretical pursuit but a practical necessity, accelerated by global crises. A pivotal turning point was the **COVID-19 pandemic**, which forced judiciaries worldwide to digitalise rapidly, effectively opening the door to broader technological innovations. This era proved that legal systems could adapt to digital environments, setting the stage for more advanced integration of AI.

Today, AI's potential in legal research is most evident when addressing the challenge of systematising massive datasets shaped by extraordinary societal events. For instance, **the ongoing war in Ukraine** has generated a vast but fragmented body of case law regarding war-related damages. With over 120 million decisions in the national archive and no dedicated filters for such unique categories, traditional manual analysis reaches its limits. Similarly, in the realm of **international human rights**, AI tools are increasingly being tested to predict judicial outcomes or identify patterns in the European Court of Human Rights' extensive jurisprudence.

In these contexts, AI offers a transformative solution. While machine-learning algorithms require expert tuning by scientists, they can process data at a scale unimaginable for humans. However, this remains a collaborative effort: legal scholars must provide the essential normative criteria and contextual datasets to ensure AI training remains grounded in the rule of law.

A compelling example of AI's potential in legal research is the challenge of systematising and correctly analysing court decisions.[8] This challenge becomes especially urgent in areas of law shaped by extraordinary societal events, such as the ongoing war in Ukraine. With more than 120 million decisions in the archive, finding relevant cases, identifying patterns, and tracing the development of this category is exceedingly difficult. AI technologies can offer a solution.[9] While AI cannot independently perform analytical work without prior training, machine-learning algorithms, developed and tuned by experts in exact sciences, can process data at a scale unimaginable for humans. Legal scholars, in turn, must provide the criteria, datasets, and contextual understanding necessary for AI training.[10]

The prospects of integrating AI into legal scholarship are substantial:

- more efficient processing of court decisions;
- improved search accuracy across massive datasets;
- the ability to extract unique cases from thousands of similar decisions;
- simplified and accelerated systematisation of judicial practice;
- timely identification of inconsistencies or gaps;
- opportunities to isolate stable judicial approaches and identify exceptional trends.

These advances can also stimulate the development of legislation grounded in real-world challenges. For example, a comprehensive analysis of case law on compensation for war-related damage may inform future compensation mechanisms and help prevent "double recovery" in judicial and extrajudicial procedures.[11]

The formation of this Special Issue is driven by the growing integration of AI tools into legal institutions and processes. Thanks to the contributions gathered here, this Special Issue brings together diverse and timely perspectives on the evolving relationship between artificial intelligence, technological development, and the law. The contributions reflect a wide geographical and thematic scope, **from Eastern Europe and Vietnam to the Middle East** and beyond.

8 Nikolaos Aletras and others, 'Predicting Judicial Decisions of the European Court of Human Rights: A Natural Language Processing Perspective' (2016) 2 PeerJ Computer Science e93. doi:10.7717/peerj-cs.93.

9 Vitaliy Golomozyi and others, 'Processing Big Data of Court Decisions' (2023) 11(4) Baltic Journal of Modern Computing 580. doi:10.22364/bjmc.2023.11.4.04.

10 Joe Collenette, Katie Atkinson, and Trevor Bench-Capon, 'Explainable AI Tools for Legal Reasoning about Cases: A Study on the European Court of Human Rights' (2023) 317(c) Artificial Intelligence 103861. doi:10.1016/j.artint.2023.103861.

11 Yuliia Hartman, 'International Experience of Damages Compensation in Armed Conflicts: Lessons for Ukraine' (2025) 14 F1000Research 1247. doi:10.12688/f1000research.171894.1.

The contributions gathered in this Special Issue reflect a broad and diverse spectrum of inquiry, united by a strong comparative dimension and an emphasis on European legal standards as a benchmark for global reform. The article from Vietnam sets this tone by offering a rigorous comparative study between the European Union and Vietnam, examining how criminal justice systems can integrate AI-driven innovations without compromising personal data protection or fundamental fair-trial rights. This theme of balancing technology with procedural integrity is further expanded in the analysis by UAE-based researchers, who investigate the growing impact of social media on criminal proceedings. By identifying legislative gaps in Jordan and proposing safeguards to protect the presumption of innocence and judicial impartiality, their work underscores the global need to adapt traditional legal principles to the digital sphere.

The European experience remains a central point of reference throughout the issue. The study from Lithuania provides a timely doctrinal analysis of the new EU Product Liability Directive, unpacking its updated framework and the critical shift toward stricter liability regimes for harm caused by AI systems. Complementing this, the contribution from Ukrainian scholars addresses the pervasive challenge of algorithmic bias. By defining the limits of technological neutrality and proposing a pragmatic governance blueprint for the rights-compliant use of large language models (LLMs) in courts, with a specific focus on the Ukrainian context, the authors bridge the gap between European theoretical standards and practical judicial application. Together, these studies illustrate the rapidly expanding intersections between AI and law, highlighting how comparative insights can help navigate the systemic challenges of digital transformation. We invite our readers to explore these timely contributions and engage with the vital questions they raise.

I would also like to express my sincere gratitude to our guest editors, **Dr. Goda Strikaitė-Latušinskaja** and **Dr. Costas Popotas**, for their exceptional dedication, meticulous work, and professionalism. Their commitment, expertise in law and AI, and thoughtful contributions to the preparation and evaluation of the manuscripts have been invaluable.

Their support greatly assisted the AJEE's editorial team throughout the development of this Special Issue, and it is no exaggeration to say that this publication would not have been possible without their efforts.

We hope that this Special Issue will contribute meaningfully to the ongoing global discussion on the expanding integration of AI into the legal sphere and the complex challenges arising from this transformation. We warmly encourage our readers to engage in post-publication dialogue and continue this increasingly important conversation.

REFERENCES

1. Aletras N and others, 'Predicting Judicial Decisions of the European Court of Human Rights: A Natural Language Processing Perspective' (2016) 2 PeerJ Computer Science e93. doi:10.7717/peerj-cs.93
2. Collenette J, Atkinson K and Bench-Capon T, 'Explainable AI Tools for Legal Reasoning about Cases: A Study on the European Court of Human Rights' (2023) 317(c) Artificial Intelligence 103861. doi:10.1016/j.artint.2023.103861
3. Golomozyi V and others, 'Processing Big Data of Court Decisions' (2023) 11(4) Baltic Journal of Modern Computing 580. doi:10.22364/bjmc.2023.11.4.04
4. Hartman Yu, 'International Experience of Damages Compensation in Armed Conflicts: Lessons for Ukraine' (2025) 14 F1000Research 1247. doi:10.12688/f1000research.171894.1

AUTHOR INFORMATION

Goda Strikaitė-Latušinskaja
PhD (Law), Junior Assistant, Faculty of Law, Vilnius University, Vilnius, Lithuania.
goda.strikaite@tf.vu.lt
https://orcid.org/0000-0003-1284-5783
Corresponding author, responsible for preparing the first section of the Editorial.

Yuliia Hartman
PhD student in Law, Taras Shevchenko National University of Kyiv, Ukraine
Managing Editor of Access to Justice in Eastern Europe Journal.
Member of EASE Ukraine.
yuliia.hartman@knu.ua
https://orcid.org/0000-0002-5637-8358
Corresponding author, responsible for preparing the second section of the Editorial. Thank our colleagues, the Editor-in-Chief, and the Managing Editors of AJEE, for their help, thoughts, and comments on my piece, which helped improve it.

Competing interests: Mag. Yuliia Hartman is the Managing Editor of AJEE and, as such, has no competing interests to declare regarding the content of this Editorial. Although Dr. Goda Strikaitė-Latušinskaja is a Guest Editor of this issue and also has no competing interests to declare regarding the content of this Editorial. This Editorial, as is standard for this publication type, was not subject to peer review but aligns fully with the Journal's publication ethics.

Disclaimer: The views and opinions expressed in this Editorial are those of the author and do not necessarily reflect the official policy or position of the Journal, the Editorial Board, the publisher, or any affiliated institutions.

RIGHTS AND PERMISSIONS

ABOUT THIS ARTICLE

Cite this article
Strikaitė-Latušinskaja G and Hartman Yu, 'From Editors: Artificial Intelligence and Law in a World of Accelerating Change' (2025) 8(Spec) Access to Justice in Eastern Europe 6-14 <https://doi.org/10.33327/AJEE-18-8.S-ed000164 >

DOI: https://doi.org/10.33327/AJEE-18-8.S-ed000164

АНОТАЦІЯ УКРАЇНСЬКОЮ МОВОЮ

СЛОВО ВІД РЕДАКТОРІВ: ШТУЧНИЙ ІНТЕЛЕКТ І ПРАВО В УМОВАХ СТРІМКИХ ГЛОБАЛЬНИХ ТРАНСФОРМАЦІЙ

АНОТАЦІЯ

Ця вступна стаття від редакторів до спеціального випуску AJEE, присвяченого штучному інтелекту та праву, розглядає сучасні технологічні розробки в контексті стрімко мінливого глобального середовища, в якому зростаюче використання цифрових та алгоритмічних інструментів перетинається з фундаментальними правами людини, публічним управлінням і ключовими функціями систем правосуддя. У тексті підкреслюється, що штучний інтелект ставить під сумнів усталені правові уявлення (зокрема щодо відповідальності, процесуальної справедливості та інституційної структури), водночас відкриваючи нові можливості для підвищення ефективності, інноваційності та доступу до правосуддя.

Матеріали цього спеціального випуску розглядають зазначені питання в різних юрисдикціях: від формування регуляторної рамки штучного інтелекту в Казахстані та складнощів застосування ШІ у судочинстві, до викликів цивільно-правового характеру в арабських правових системах, розвитку кримінального правосуддя в Німеччині та Україні, а також еволюції управління цифровими правами в Саудівській Аравії. У сукупності ці дослідження демонструють, що правові й інституційні питання, породжені застосуванням штучного інтелекту, не можуть бути обмежені рамками однієї дисципліни чи національного контексту; натомість вони потребують міждисциплінарного, порівняльного та орієнтованого на права людини підходу, чутливого як до локальних реалій, так і до глобальних стандартів. У вступі запропоновано запрошення для науковців і практиків замислитися над умовами, за яких штучний інтелект може посилювати, а не підривати, верховенство права та суспільну довіру.

Access to Justice in Eastern Europe
ISSN 2663-0575 (Print)
ISSN 2663-0583 (Online)
http://ajee-journal.com

RESEARCH ARTICLE

Research Article

AI SYSTEMS IN JUDICIAL DECISION-MAKING: SUPPORT VS. SUPERSEDE — THE EUROPEAN PERSPECTIVE

Goda Strikaitė-Latušinskaja

DOI:
https://doi.org/10.33327/AJEE-18-8.S-c000163

Date of submission: 28 Oct 2025
Date of acceptance: 18 Dec 2025
Date of Publication: 30 Dec 2025

Disclaimer:
The author declares that their opinion and views expressed in this manuscript are free of any impact of any organizations.

ABSTRACT

Background: *As technological progress accelerates within the judiciary, debate is intensifying over whether artificial intelligence (AI) could, or should, replace human judges in the decision-making process. Increasing attention is being paid to the possibility that AI systems may, over time, equal or surpass human judges in efficiency, consistency, and the delivery of reasoned decisions. At the same time, current developments in legal technology primarily point toward the use of AI as a tool designed to assist judicial decision-making rather than to exercise autonomous adjudicatory authority. This tension between supportive and substitutive uses of AI highlights the need for a nuanced analysis of the permissible and appropriate role of AI in adjudication.*

The debate becomes even more complex in the European context, where the intersection of technology and law is guided by a commitment to upholding fundamental rights and ethical principles. The adoption of various soft law instruments, such as ethical guidelines and recommendations on AI, alongside the binding provisions of the Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down

harmonised rules on artificial intelligence (the AI Act), underscores the EU's proactive approach to regulating AI in high-risk and sensitive domains, including the administration of justice. This dual emphasis on ethical standards and legal safeguards makes it essential to examine the European approach to AI in adjudication.

Methods: *This article employs a qualitative legal methodology, drawing primarily on doctrinal, analytical, and teleological methods. The doctrinal method serves as the foundation, involving a systematic analysis of EU and Council of Europe instruments, including the European Ethical Charter on the Use of AI in Judicial Systems, the Ethics Guidelines for Trustworthy AI, and the AI Act, to identify how European law conceptualises AI in adjudication and safeguards human oversight. The teleological method is applied to interpret these instruments in light of their broader objectives, uncovering how human-centric principles and fundamental rights guide the permissible use of AI in courts. Finally, the analytical method integrates insights from these sources to develop a conceptual framework distinguishing between supportive and substitutive models of AI adjudication, thereby clarifying the normative boundaries of the European approach.*

Results and conclusions: *The paper concludes that the European approach to AI in adjudication is defined by a human-centric and rights-based paradigm, developed through the combined efforts of the European Union and the Council of Europe. The results show that this framework consistently positions AI as a supportive tool that enhances judicial efficiency and consistency, while ensuring that final decision-making authority remains with human judges. At the same time, the analysis recognises that this approach, though coherent and well-suited to current technological realities, may increasingly be tested as AI systems become more advanced, challenging the assumptions that underpin the current European model, built on human oversight and control. While the framework firmly excludes autonomous AI judges, future developments may prompt renewed consideration of whether its existing boundaries remain adequate to govern increasingly sophisticated technological involvement in adjudication.*

1 INTRODUCTION

The extent and intensity of technology use in courts vary significantly across regions and countries. Undoubtedly, one of the main factors behind such uneven an distribution of technology is legal regulation. The United States, China, and the European Union are competing to become the global leader in finding the right balance between effective regulation and the widespread application of digital technologies. In general, there are three different competing regulatory approaches: the American market-driven model, the Chinese state-driven model, and the European rights-driven regulatory model.[1]

1 See more about the three digital empires and their competing models in Anu Bradford, *Digital Empires: The Global Battle to Regulate Technology* (OUP 2023).

Artificial Intelligence (AI) is deemed the key enabling technology,[2] and with the European Union's ambition to position itself as a global leader in developing cutting-edge, trustworthy AI,[3] it raises important questions: What types of AI tools are suitable for adoption in European countries? Specifically, in the context of courts, to what extent can AI be integrated into judicial processes? Critically, can AI be applied to the core of judges' work, decision-making? Could it fully take over the adjudication role?

In this context, two distinct models of applying AI in courts can be observed. The first model uses AI as a decision-support tool to enhance judicial efficiency, accuracy, and consistency while retaining final authority in human judges. The second model envisages AI as an autonomous decision-maker, or so-called robot judge, capable of conducting proceedings and issuing binding decisions with minimal or no human involvement. These contrasting approaches, supportive versus substitutive, form the analytical framework of this article.

Accordingly, the article explores the role of AI in judicial decision-making, focusing on identifying the distinctive European approach to this issue. To do that, the article is structured as follows. Section 2 examines the core aspects of the European approach to AI in adjudication, emphasising the balance between AI support and human oversight, adherence to ethical principles, implementation of safeguards, and analysis of key policy documents. Section 3 analyses the potential for AI to assist judges in the decision-making process while preserving human oversight, whereas Section 4 delves into the prospects of AI replacing judges in decision-making altogether. Finally, Section 5 concludes by summarising the article's main findings.

2 EUROPEAN APPROACH TO AI IN ADJUDICATION

The European approach to AI in adjudication has developed through the combined efforts of both the Council of Europe and the institutions of the European Union, united by a commitment to ensure that technology serves rather than supplants human judgment. In 2018, the Council of Europe's European Commission for the Efficiency of Justice (CEPEJ) recognised AI as one of the technologies with the greatest potential to enhance the efficiency and quality of justice, while warning against its uncritical use for predicting or automating judicial decisions.[4] That same year, the CEPEJ adopted the European Ethical Charter on the

2 The 2018 report of the High Level Group on Industrial Technologies recognised AI as a "key enabling technology", highlighting the transformative role of AI and the necessity for the industry to use AI to maintain its leadership, see: European Commission, *Re-finding Industry – Defining Innovation: Report of the independent High Level Group on industrial technologies* (Publications Office of the EU 2018) 20. doi:10.2777/927953.

3 See, for example, High-Level Expert Group on AI, *Ethics Guidelines for Trustworthy AI* (European Commission 2019).

4 CEPEJ, *European Judicial Systems: Efficiency and Quality of Justice* (CEPEJ studies no 26, edn 2018: data 2016) <https://book.coe.int/en/international-law/7698-european-judicial-systems-2018-edition-2016-data-efficiency-and-quality-of-justice.html> accessed 30 September 2025.

Use of AI in Judicial Systems, which affirmed that the deployment of AI in courts should improve efficiency and consistency but must always respect fundamental rights, judicial independence, and the right to a fair trial.[5] Parallel discussions were unfolding within the European Union: following a 2017 request by the European Council to develop a coordinated European approach to AI,[6] the European Commission's 2018 Communication AI for Europe identified justice as a key application area and introduced a dual approach, promoting innovation while ensuring adherence to the Union's values and ethical principles of accountability and transparency.[7] Subsequent EU policy documents, including the 2019–2023 European e-Justice Strategy and Action Plan,[8] further distinguished AI as a promising but high-risk technology, whose development must be accompanied by safeguards in data protection, ethics, and fundamental-rights compliance. Collectively, these initiatives reflect a distinctly European approach to AI in adjudication, one that encourages technological innovation while maintaining the primacy of human oversight, judicial independence, and fundamental rights.

Building on these initiatives, European institutions articulated a more detailed ethical and governance framework for the use of AI in the justice system. Between 2019 and 2021, a series of soft-law instruments were adopted by the European Commission, the European Parliament, the Council of the European Union, and the Council of Europe, each reinforcing the principle that AI technologies should support judicial activity without undermining human oversight or judicial independence.

In 2019, the High-Level Expert Group on AI, established by the European Commission, presented the Ethics Guidelines for Trustworthy AI, marking a key step in shaping the European vision of human-centric AI. Among the seven requirements for trustworthy AI identified in the guidelines, particular importance is placed on human agency, oversight, and respect for human dignity. In the context of adjudication, these principles mean that AI systems should act as tools serving and protecting human autonomy rather than replacing it, and that final decisions must remain under meaningful human control through mechanisms such as human-in-the-loop or human-in-command oversight.[9] The same year, the European Commission issued the Communication Building Trust in

5 CEPEJ European Ethical Charter on the Use of Artificial Intelligence (AI) in Judicial Systems and their Environment (adopted 3-4 December 2018) <https://www.coe.int/en/web/cepej/cepej-european-ethical-charter-on-the-use-of-artificial-intelligence-ai-in-judicial-systems-and-their-environment> accessed 30 September 2025.

6 Laura Delponte, *European Artificial Intelligence (AI) Leadership, the Path for an Integrated Vision* (European Parliament 2018) <https://www.europarl.europa.eu/thinktank/en/document/IPOL_STU(2018)626074> accessed 30 September 2025.

7 European Commission, *Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions Artificial Intelligence for Europe* (COM/2018/237 final, 25 April 2018) <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52018DC0237> accessed 30 September 2025.

8 Council of the EU, *2019–2023 Action Plan European e-Justice* ST/5140/2019/INIT [2019] OJ C 96/9.

9 High-Level Expert Group on AI (n 3).

Human-Centric AI, reaffirming that AI systems should support individuals in making better, more informed choices in accordance with their goals. They should function as enablers of a flourishing and equitable society by strengthening human agency and fundamental rights, and should not decrease, limit, or misguide human autonomy. The Communication further stressed that human oversight helps ensure that an AI system does not undermine human autonomy or cause other adverse effects.[10] In the White Paper on AI (2020), the Commission noted that the appropriate type and degree of human oversight may vary from one case to another, depending on the intended use of the systems and the effects that the use could have for affected citizens and legal entities.[11] This reflects the flexible yet indispensable role of human supervision, which is particularly relevant in the context of adjudication. Building on this understanding, the Commission's 2020 Communication Digitalisation of Justice in the European Union: A Toolbox of Opportunities applied these principles directly to the justice sector. The document emphasised that final decision-making must remain a human-driven process. Only a judge can guarantee genuine respect for fundamental rights, balance conflicting interests and reflect the constant changes in society in the analysis of a case.[12]

The European Parliament also echoed this human-centric approach in its Resolution on a Framework of Ethical Aspects of AI, Robotics and Related Technologies, adopted in October 2020. The Resolution emphasised that technologies capable of making automated decisions and thereby altering the determinations of public authorities should be approached with great caution, particularly in the field of justice. It further underlined that member states should use such technologies only if there is detailed evidence of their reliability and if a meaningful human review is possible.[13] Through this, the Parliament reaffirmed that technological innovation must not come at the expense of fundamental rights or human autonomy. The Council of the European Union adopted a similar position in its Conclusions of 8 October 2020, titled "Access to Justice – Seizing the Opportunities of

10 European Commission, *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions Building Trust in Human-Centric Artificial Intelligence* (COM/2019/168 final, 8 April 2018) <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52019DC0168> accessed 30 September 2025.

11 European Commission, *White Paper: On Artificial Intelligence – A European Approach to Excellence and Trust* (COM(2020) 65 final, 19 February 2020) <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52020DC0065> accessed 30 September 2025.

12 European Commission, *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and The Committee of the Regions: Digitalisation of Justice in the European Union A Toolbox of Opportunities* (COM/2020/710 final, 2 December 2020) <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=COM:2020:710:FIN> accessed 30 September 2025.

13 European Parliament, *Report with Recommendations to the Commission on a Framework of Ethical Aspects of Artificial Intelligence, Robotics and Related Technologies* (A9-0186/2020, 2020/2012(INL), 8 October 2020) <https://www.europarl.europa.eu/doceo/document/A-9-2020-0186_EN.html> accessed 30 September 2025.

Digitalisation". The Council stressed that the use of AI tools must not interfere with judges' decision-making power or judicial independence. A court decision must always be made by a human being and cannot be delegated to an AI tool.[14]

Recent developments at the Council of Europe level further reinforce this human-centred approach. In September 2024, the Council of Europe adopted the Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law,[15] the first binding international instrument addressing the lifecycle of artificial intelligence systems used by public authorities. The Convention establishes horizontal obligations requiring that activities involving AI remain fully consistent with human rights, democracy, and the rule of law, and explicitly calls on member states to ensure that AI systems are not used in a manner that undermines judicial independence, the separation of powers or access to justice. It further requires the adoption of measures ensuring transparency, oversight, accountability and responsibility for adverse impacts on fundamental rights, as well as context and risk-based safeguards, documentation of AI systems capable of significantly affecting human rights, and the availability of effective remedies, including the ability to contest decisions substantially informed by AI use. While not regulating adjudication as such, these requirements situate the use of artificial intelligence in the justice domain within a governance framework that presupposes human authority, responsibility and procedural fairness in the exercise of judicial power, thereby limiting AI to a supportive role rather than an autonomous one.

These soft-law instruments paved the way for the European Union's next step: the adoption of the AI Act,[16] the world's first binding legal framework on AI, which anchors Europe's human-centric approach and emphasises that AI should support rather than replace human judgement.

By classifying as high-risk those AI systems intended to be used by a judicial authority or on their behalf to assist a judicial authority in researching and interpreting facts and the law and in applying the law to a concrete set of facts,[17] the EU recognises both the potential and the sensitivity of such tools. Given that high-risk AI systems must comply with the additional safeguards set out in Section 2 of the AI Act, including, *inter* alia, the requirement of effective human oversight under Article 14. The AI Act envisages only

14 Council of the EU, *Council Conclusions 'Access to Justice – Seizing the Opportunities of Digitalisation'* 2020/C 342 [2020] OJ C 342 I/1.

15 Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (5 September 2024) [2024] CETS 225.

16 Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying down Harmonised Rules on Artificial Intelligence and Amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act – AI Act) [2024] OJ L 2024/1689 <http://data.europa.eu/eli/reg/2024/1689/oj> accessed 30 September 2025.

17 ibid, annex III, para 8(a).

supportive, not autonomous, uses of AI in adjudication. Human oversight must be designed into every high-risk system to ensure that natural persons can properly understand the system's capacities and limitations, monitor its functioning, and intervene where necessary, including by overriding or reversing the system's output. The AI Act further warns against automation bias and explicitly requires that no decision be made solely on the basis of AI-generated output without separate human verification. Taken together, these safeguards affirm that AI may support judicial reasoning, but adjudication itself remains an inherently human prerogative.

This position is explicitly reinforced in Recital 61 of the AI Act, which classifies as high-risk those AI systems intended to be used by or on behalf of a judicial authority to assist in researching and interpreting facts and the law and in applying the law to a concrete set of facts. The Recital states that such systems should be treated as high-risk given their potentially significant impact on democracy, the rule of law, individual freedoms, and the right to an effective remedy and to a fair trial. It makes clear that while AI tools can support judges' decision-making power or judicial independence, they should not replace it: the final decision-making must remain a human-driven activity.

Taken together, these instruments articulate a coherent European vision of AI in adjudication, one that welcomes technological progress while safeguarding the primacy of human judgement. The European framework views AI as a means to enhance the efficiency, consistency, and accessibility of justice, yet insists that the authority to decide must remain with human judges.

This approach is consistent with the European e-Justice Strategy 2024–2028,[18] adopted in January 2025, which situates artificial intelligence within the broader digital transformation of justice, and identifies AI as a technology capable of supporting the work of courts and justice professionals, while repeatedly emphasising the need to respect judicial independence, the rule of law, and fair trial guarantees. Notably, the Strategy refers to AI only in relation to clearly supportive functions, such as data analysis, transcription, anonymisation of judicial decisions and legal research, without envisaging its use as an autonomous decision-maker. This sets the stage for the following analysis, which examines how AI systems can, in practice, support judicial decision-making without compromising independence or accountability.

18 Council of the EU, *European e-Justice Strategy 2024-2028* ST/15509/2023/INIT [2025] OJ C 2025/437 <http://data.europa.eu/eli/C/2025/437/oj> accessed 30 September 2025.

3 AI AS JUDICIAL ASSISTANTS: SUPPORTING HUMAN DECISION-MAKING

In the European Commission's study on the use of innovative technologies in the justice field, it is indicated that, among the good practices currently in place in member states, there are already those that concern areas such as, *inter alia*, anonymisation of documents (for example, court decisions); speech-to-text and transcription; introduction of chatbots for strengthening the access to justice and public services, and Robot Process Automation for increasing efficiency and minimizing errors in repetitive tasks.[19]

In addition, the European Council notes that AI systems in the justice sector may, in the future, be capable of performing increasingly complex tasks, such as analysing, structuring, and preparing information on the subject matter of cases, automatically transcribing records of oral hearings, offering machine translation, supporting the analysis and evaluation of legal documents and court/tribunal judgements, estimating the chances of success of a lawsuit, automatically anonymising case law, and providing information via legal chatbots.[20] This raises the question: *Could AI systems one day replace judges in their core function, i.e., decision-making?* Under the current European legal framework, the answer is *no*. The EU law explicitly classifies AI systems intended to support judges as high-risk, thereby ensuring that AI may only operate in an assistive capacity. In Sourdin's words, assistant 'co-bots' rather than replacement robot judges could play a more important role in the future.[21] Furthermore, Zeleznikow noted that while robots are unlikely to replace judges, automated tools are being developed to support legal decision- making.[22] Even considering that AI has the potential to surpass judges in adjudication, AI systems complementing judicial work, including in the decision-making process, are more feasible in the near future.

AI systems can support judges by making predictions about how a case should be decided, as well as by generating a draft judgment based on those predictions. Accordingly, a human judge retains the discretion to make on the final decision and bears responsibility for that judgment. For example, Harvey, a generative AI platform developed for legal professionals and used primarily in legal practice rather than in courts, offers an illustrative account of how AI systems could, in principle, support legal reasoning without

19 Miglena Vucheva, Margarida Rocha and Robrecht Renard, *Study on the Use of Innovative Technologies in the Justice Field: Final Report* (Publications Office of EU 2020). doi:10.2838/585101.

20 Council of the EU, Council Conclusions 'Access to Justice – Seizing the Opportunities of Digitalisation' 2020/C 342 I/01 [2020] OJ C 342 I/1.

21 Tania Sourdin and Richard Cornes, 'Do Judges Need to Be Human? The Implications of Technology for Responsive Judging' in Tania Sourdin and Archie Zariski (eds), *The Responsive Judge: International Perspectives* (Ius Gentium: Comparative Perspectives on Law and Justice, Springer 2018) 87. doi:10.1007/978-981-13-1023-2_4.

22 John Zeleznikow, 'Can Artificial Intelligence and Online Dispute Resolution Enhance Efficiency and Effectiveness in Courts' (2017) 8(2) International Journal for Court Administration 30. doi:10.18352/ijca.223.

displacing human decision-making. There are databases that use natural language processing to assist with sourcing relevant material based on search terms. The system would need to go beyond these databases by reducing the returned sources to a manageable, relevant sample, then deploying tools to compare these sources of law with a present case and engaging in analysis to decide the outcome. Harvey explains that this final step requires "the development of the necessary algorithms that could undertake the comparative and predictive analysis, together with a form of probability analysis to generate an outcome that would be useful and informative."[23] However, Harvey's model retains the principle of human judge decision-making.[24]

By combining the ability of predictive systems to identify patterns that influence projections, and the ability to generate a decision on top of the predictions, in accordance with a specific case and based on the information input, AI assistants can serve judges in adjudication. A human judge could then use this draft (since many judges, especially in appellate courts, already use drafts prepared by their legal assistants) to draft their own reasons for judgment. This use of AI would allow human oversight of the computer program and enable a human judge to take into account discretionary or social considerations that may be beyond computer program's capacity or authority.[25] Much like legal clerks, AI systems can assist in preparing drafts, leaving the judge to exercise judgement, incorporate social and moral considerations, and ensure the final decision aligns with legal and constitutional principles.

The rationale for adopting such supportive systems in Europe is clear: it reflects the EU's consistent policy that AI should assist, not replace, human judges. The European approach provides clear normative boundaries for such supportive uses. As established in the preceding section, the principle of human oversight, embedded in multiple EU instruments and codified in the Artificial Intelligence Act, requires that judicial decision-making remain under human control. By classifying as high-risk those AI systems intended to assist judges in interpreting facts and law, the EU recognises their potential value while drawing a firm line against their autonomous use in adjudication.

Critically, the principle of human oversight ensures that judges retain responsibility for the final decision, preventing accountability gaps and reinforcing public trust in the judiciary. This approach also aligns with the constitutional principle of judicial prerogative. For instance, under Article 109 of the Constitution of the Republic of Lithuania, justice is administered exclusively by judges,[26] a mandate that underscores why human responsibility cannot be outsourced to machines.

23 David Harvey, 'From Susskind to Briggs: Online Court Approaches' (2016) 5 Journal of Civil Litigation and Practice 94.

24 Tania Sourdin, 'Judge v Robot? Artificial Intelligence and Judicial Decision-Making' (2018) 41(4) University of New South Wales Law Journal 1114. doi:10.3316/agis.20190207006418.

25 Sourdin and Cornes (n 21).

26 Constitution of the Republic of Lithuania (adopted 25 October 1992) <https://www.e-tar.lt/portal/tt/legalAct/TAR.47BB952431DA/ZQmhQugYfg> accessed 30 September 2025.

Recent European judicial practice illustrates both the possibilities and the risks associated with judges' use of generative AI as a supportive tool. In a judgment of 7 June 2024, the Dutch Rechtbank Gelderland openly acknowledged its use of ChatGPT to obtain information relevant for the assessment of damages, including estimates concerning the lifespan of solar panels and electricity prices.[27] While the court treated ChatGPT as a source of background information rather than a decision-maker, subsequent scholarly commentary has critically highlighted the dangers of relying on generative AI without sufficient transparency, verifiability, and methodological clarity, particularly given the risk of hallucinations and the absence of identifiable sources.[28] This example demonstrates that even limited, supportive uses of AI by judges raise serious concerns about accountability, contestability, and the right to a fair trial, underscoring the importance of the safeguards discussed in Section 2, including effective human oversight and transparency requirements.

To sum up, the use of AI systems to assist judges in preparing judgements is in line with the current European approach to preserve the human oversight in the adjudication process, which would contribute to avoiding accountability gaps and preserving public trust, as well as being consistent with the courts' fundamental role in administering justice, preserving judicial authority, and maintaining public trust, as exemplified by the constitutional principles and the importance of human responsibility in decision-making.

It should be noted that, given the state-of-the-art level of technology, examples from outside the European Union show that the implementation of AI (understood in a broad sense to include algorithmic and data-driven systems supporting judicial decision-making) already raises serious concerns about access to justice. In the United States, the COMPAS algorithm, used to assess the likelihood of recidivism, has been criticised for racial bias and a lack of transparency,[29] undermining the fairness and contestability of judicial decisions. In China, the Smart Courts system goes further by issuing "abnormal judgment warnings" when rulings deviate from prior cases,[30] thereby pressuring judges to conform to algorithmic patterns rather than exercising independent judicial reasoning. These experiences demonstrate that even supportive, non-autonomous systems, designed to aid rather than replace judges, may inadvertently compromise fundamental guarantees of access to justice if their operation is opaque, biased, or insufficiently overseen by humans.

27 ECLI:NL:RBGEL:2024:3636 (Gelderland District Court, 7 June 2024) <https://www.recht.nl/rechtspraak/?ecli=ECLI:NL:RBGEL:2024:3636> accessed 30 September 2025.

28 André Janssen, 'Editorial: The Use of ChatGPT by the Judge: What Can Go Wrong, Goes Wrong' (2024) 32(5) European Review of Private Law 741.

29 Julia Angwin and others, 'Machine Bias' (*ProPublica*, 23 May 2016) <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> accessed 30 September 2025.

30 Brian M Barry, *How Judges Judge: Empirical Insights into Judicial Decision-Making* (Routledge 2023).

4 AI AS AUTONOMOUS ADJUDICATORS: THE PROSPECT OF REPLACING JUDGES

What concerns the use of AI to replace judges in adjudication, according to Volokh, is that if an AI program someday passes a Turing test, and its developers can then teach it to converse, and even present an extended persuasive argument. Moreover, if the software can create persuasive opinions, capable of regularly winning opinion-writing competitions against human judges, we should, in principle, accept it as a judge.[31] Agreeing with Mizaras, and in light of the European Court of Human Rights interpretation of the Convention as a 'living instrument' that evolves alongside societal developments, we cannot entirely rule out the possibility of an autonomous E-Judge in the future. While still a work in progress, such a judge could potentially be tasked with resolving smaller, repetitive cases of a non-complex nature, i.e., cases characterised by clear outcomes and minimal evidentiary or interpretive challenges. Realising this vision, however, would require society to approach AI with an open mind, while upholding rigorous legal and ethical standards.[32]

Even though the current state-of-the-art AI, and the discussed European approach in terms of using this technology in courts, suggest that we should still be reserved in talking about autonomous application of AI in adjudication, in R. Susskind's words, lawyers, judges, and policy- makers should be both humbled and open-minded about as-yet-uninvented technologies.[33] A few considerations to bear in mind, supporting the idea that in the not-so-near future, the question may not be whether AI will be capable of replacing judges, but whether societies outside the European model will be willing to accept such a transformation, and what implications this might have for global standards of justice, will be discussed below.

Firstly, when considering the limits of the application of AI in courts, what is often underestimated is the so-called AI fallacy, i.e., the view that systems cannot replicate human lawyers and their judgement because they cannot, *inter alia*, exercise judgement or be empathetic. However, the problem with this point of view is thinking that the *only* way to get machines to outperform humans is to mimic human reasoning and the way humans work. However, the present-day AI systems operate do not by copying human beings, they take on the work in ways that are best suited to their unique capabilities, not ours.[34] The second wave of artificial intelligence has provided an answer to sceptical arguments, after maintaining for a long time that AI could never replace a human, because it does not have and will never have the human qualities. However, the new approach to reach the aimed

31 Eugene Volokh, *'Chief Justice Robots' (2019)* 68(6) Duke Law Journal 1135.

32 Vytautas Mizaras, 'Artificial Intelligence and the Right to a Fair Trial' (Opening of the Judicial Year 2025: Judicial Seminar, ECHR, 31 January 2025) <https://www.echr.coe.int/documents/d/echr/speech-20250131-mizaras-jy-eng> accessed 30 September 2025.

33 Richard Susskind, *Online Courts and the Future of Justice* (OUP 2019).

34 ibid, see more on AI fallacy.

result in the most feasible way, rather than to mimic human-beings, opened new possibilities in terms of the ways AI could be used. It is not to mention that there will be a third, as well as a fourth, wave of artificial intelligence, promising even more sophisticated results, thus suggesting that AI could excel in judicial decision-making. As Susskind famously remarked, "patients do not want neurosurgeons; they want health."[35]

Secondly, as systems are improving and taking-over more and more tasks, which had previously been exclusively attributed to human abilities, also, as they, *inter alia*, are making more accurate predictions[36] and more proficient at playing chess,[37] it is likely that technological intrusion we are comfortable with delegating tasks will change to a great extent. In this era of increasingly capable machines, then, it is not outrageous to expect at some stage, whether twenty or one hundred years from now, that systems will outperform judges at their own game, by delivering reasoned judgements with explanations that will look and feel like the finest of human judgements but sourced through AI rather than the judicial 'wetware'.[38] As technology progresses, society's comfort with delegating tasks to AI will grow, and it is plausible that AI could deliver judgements even surpassing those of human judges.

From a European regulatory perspective, however, the prospect of autonomous AI adjudication remains incompatible with the existing legal framework. As discussed in Section 2, the AI Act classifies AI systems used to assist judicial authorities as high-risk and subjects them to strict human oversight obligations, while Recital 61 explicitly presupposes that adjudication remains a human-driven activity. Together with the Council of Europe's insistence on judicial independence, accountability and effective remedies, this framework leaves little room for the lawful deployment of fully autonomous adjudicators in Europe under current conditions. Any move towards replacing judges with AI would require not only technological advances but also a fundamental reconfiguration of Europe's constitutional and human rights-based understanding of adjudication.

To conclude, this evolution, i.e., AI systems analysing facts, applying the law more effectively than humans, and even delivering reasoned judgements that would rival or surpass those written by flesh-and-blood judges, would challenge the foundations of the European model of adjudication, which rests on human oversight as an indispensable precondition for the proper upholding of the principle of access to justice. Decision-makers and stakeholders would need to reconsider the current boundaries of AI

35 ibid, 286.

36 For example, *Lex Machina* is deemed to make better predictions than lawyers, see more: 'Lex Machina® Actionable Intelligence: Now empowered by LexisNexis Protégé™' (*LexisNexis,* 2025) <https://lexmachina.com/legal-analytics/> accessed 30 September 2025.

37 For example, as early as 1997, the incumbent world chess champion, Garry Kasparov, was beaten by IBM's *Deep Blue* chess-playing expert system: Feng-Hsiung Hsu, *Behind Deep Blue: Building the Computer That Defeated the World Chess Champion* (Princeton UP 2002).

38 Susskind (n 33).

applications in justice systems and potentially modify the existing legal framework to meet the demands of emerging realities while safeguarding enduring values and fundamental principles. While these claims remain speculative, they highlight that, in the long term, society may face not only the question of AI's capabilities but also the question of whether we are comfortable delegating adjudication to a non-human system.

5 CONCLUSIONS

The analysis demonstrates that the European approach to AI in adjudication is firmly grounded in a human-centric, rights-based framework. The Council of Europe and the European Union both view AI as a tool to enhance the efficiency, consistency, and accessibility of justice, not as a substitute for judicial reasoning or human judgment. The AI Act and accompanying soft-law instruments establish a clear normative boundary: AI may assist judges in interpreting facts and law, yet the final decision must remain human-driven and subject to meaningful oversight.

For now, to ensure procedural fairness, judicial accountability, and public confidence, the role of AI in adjudication must remain limited to a supportive function, with judges retaining full authority and responsibility for their decisions. Properly implemented, AI systems can improve efficiency and consistency without undermining judicial independence or the constitutional mandate to administer justice.

At the same time, continued technological development may prompt a careful reassessment of how these safeguards operate in practice. The current European framework, developed through sustained collaboration between EU institutions and the Council of Europe, remains appropriate for present capabilities. Yet future advancements could necessitate a nuanced recalibration of the existing approach to ensure that its commitment to human oversight remains both principled and effective. The central issue will not be AI's technical sophistication, but how far societies are prepared to rely on non-human systems without compromising the foundations of justice.

REFERENCES

1. Angwin J and others, 'Machine Bias' (*ProPublica*, 23 May 2016) <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> accessed 30 September 2025
2. Barry BM, *How Judges Judge: Empirical Insights into Judicial Decision-Making* (Routledge 2023)
3. Bradford A, *Digital Empires: The Global Battle to Regulate Technology* (OUP 2023)
4. Delponte L, *European Artificial Intelligence (AI) Leadership, the Path for an Integrated Vision* (European Parliament 2018)

5. Harvey D, 'From Susskind to Briggs: Online Court Approaches' (2016) 5 Journal of Civil Litigation and Practice 84
6. Hsu FH, *Behind Deep Blue: Building the Computer That Defeated the World Chess Champion* (Princeton UP 2002)
7. Janssen A, 'Editorial: The Use of ChatGPT by the Judge: What Can Go Wrong, Goes Wrong' (2024) 32(5) European Review of Private Law 741
8. Mizaras V, 'Artificial Intelligence and the Right to a Fair Trial' (Opening of the Judicial Year 2025: Judicial Seminar, ECHR, 31 January 2025)
9. Sourdin T and Cornes R, 'Do Judges Need to Be Human? The Implications of Technology for Responsive Judging' in Sourdin T and Zariski A (eds), *The Responsive Judge: International Perspectives* (Ius Gentium: Comparative Perspectives on Law and Justice, Springer 2018) 87. doi:10.1007/978-981-13-1023-2_4
10. Sourdin T, 'Judge v Robot? Artificial Intelligence and Judicial Decision-Making' (2018) 41(4) University of New South Wales Law Journal 1114. doi:10.3316/agis.20190207006418
11. Susskind R, *Online Courts and the Future of Justice* (OUP 2019)
12. Volokh E, *'Chief Justice Robots' (2019)* 68(6) Duke Law Journal 1135
13. Vucheva M, Rocha M and Renard R, *Study on the Use of Innovative Technologies in the Justice Field: Final Report* (Publications Office of EU 2020). doi:10.2838/585101
14. Zeleznikow J, 'Can Artificial Intelligence and Online Dispute Resolution Enhance Efficiency and Effectiveness in Courts' (2017) 8(2) International Journal for Court Administration 30. doi:10.18352/ijca.223

AUTHORS INFORMATION

Goda Strikaitė-Latušinskaja
PhD (Law), Junior Assistant, Faculty of Law, Vilnius University, Vilnius, Lithuania
goda.strikaite@tf.vu.lt
https://orcid.org/0000-0003-1284-5783
Corresponding author, solely responsible for the manuscript preparing.

Competing interests: Although Dr. Goda Strikaitė-Latušinskaja is a Guest Editor of this issue, the article underwent independent peer review, and Dr. Strikaitė-Latušinskaja's involvement in the editorial process did not in any way influence the decision regarding the acceptance of this article.

Disclaimer: The author declares that their opinion and views expressed in this manuscript are free of any impact of any organizations.

RIGHTS AND PERMISSIONS

EDITORS

ABOUT THIS ARTICLE

Cite this article
Strikaitė-Latušinskaja G, 'AI Systems in Judicial Decision-Making: Support vs Supersede — The European Perspective' (2025) 8(Spec) Access to Justice in Eastern Europe 15-31 <https://doi.org/10.33327/AJEE-18-8.S-c000163>

DOI: https://doi.org/10.33327/AJEE-18-8.S-c000163

Keywords: *courts; adjudication; artificial intelligence; robot-judge; European regulation; automated decision-making; AI decision-support systems.*

DETAILS FOR PUBLICATION

Date of submission: 28 Oct 2025
Date of acceptance: 18 Dec 2025
Last publication: 30 Dec 2025
Whether the manuscript was fast tracked? - No
Number of reviewer report submitted in first round: 2 reports
Number of revision rounds: 1 round with conditional acceptance

Technical tools were used in the editorial process:
Plagiarism checks - Turnitin from iThenticate https://www.turnitin.com/products/ithenticate/
Scholastica for Peer Review https://scholasticahq.com/law-reviews

AI DISCLOSURE STATEMENT

The author confirms that the manuscript was prepared by the authors. AI tools were employed exclusively for spelling, grammar, and stylistic refinement. No generative AI was used to produce original content, research ideas, or analysis.

АНОТАЦІЯ УКРАЇНСЬКОЮ МОВОЮ

Дослідницька стаття

СИСТЕМИ ШІ ПІД ЧАС УХВАЛЕННЯ СУДОВИХ РІШЕНЬ: ПІДТРИМКА ПРОТИ ЗАМІЩЕННЯ — ЄВРОПЕЙСЬКА ПЕРСПЕКТИВА

Ґода Стрікайте-Латушинська

АНОТАЦІЯ

Вступ: *З прискоренням технологічного прогресу, зокрема в судовій системі, загострюються дискусії щодо того, чи може або чи повинен штучний інтелект (ШІ) замінити суддів-людей у процесі ухвалення рішень. Дехто підтримує цю ідею, зазначаючи, що, наприклад, судді-ШІ будуть не менш надійними (і ефективнішими), ніж судді-люди. І якщо програма ШІ колись пройде тест Тюрінга, ми, в принципі, повинні прийняти її як суддю, або що в цю епоху новітніх технологій не є неприпустимим очікувати на певному етапі, чи то через двадцять, чи через сто років, що системи перевершать суддів у їхній власній справі. Крім того, в епоху дедалі потужніших машин цілком можливо, що системи ШІ зрештою можуть перевершити суддів в ухваленні обґрунтованих та послідовних рішень. З іншого боку, критики вважають, що хоча роботи навряд чи замінять суддів, розробляються автоматизовані інструменти для підтримки ухвалення юридичних рішень, і що помічники-«коботи», а не роботи, що заміняють суддів, відіграватимуть більш значну роль у майбутньому. Для вивчення можливостей використання штучного інтелекту для підтримки або заміщення суддів у процесі ухвалення рішень необхідний нюансований аналіз.*

Дебати стають ще складнішими в європейському контексті, де перетин технологій та права керується зобов'язанням дотримуватися основних прав та етичних принципів. Прийняття різних інструментів «м'якого» права, таких як етичні рекомендації та рекомендації щодо штучного інтелекту, поряд із обов'язковими положеннями Акту про ШІ, підкреслює проактивний підхід ЄС до регулювання ШІ у сферах високого ризику та чутливості, зокрема щодо здійснення правосуддя. Цей подвійний акцент на етичних стандартах та правових гарантіях робить необхідним вивчення європейського підходу до ШІ у судочинстві.

Методи: *У цій статті використовується якісна правова методологія, що спирається переважно на доктринальні, аналітичні та телеологічні методи. Доктринальний метод слугує основою, що містить систематичний аналіз інструментів ЄС та Ради Європи, зокрема Європейської етичної хартії щодо використання штучного інтелекту в судових системах, Етичних рекомендацій щодо надійного штучного інтелекту та Акту про ШІ, щоб визначити, як європейське законодавство концептуалізує штучний інтелект у судовому розгляді та гарантує людський нагляд. Телеологічний метод застосовується для інтерпретації цих інструментів у світлі їхніх ширших цілей, розкриваючи, як людиноцентричні принципи та основні права керують допустимим використанням штучного інтелекту в судах. Нарешті, аналітичний метод інтегрує знання з цих джерел для розробки концептуальної основи, яка розрізняє моделі підтримки та заміщення у судовому розгляді на основі штучного інтелекту, тим самим уточнюючи нормативні межі європейського підходу.*

Результати та висновки: *У статті було зроблено висновок, що європейський підхід до штучного інтелекту в судовому розгляді визначається людиноцентричною та правоорієнтованою парадигмою, розробленою завдяки спільним зусиллям Європейського Союзу та Ради Європи. Результати показують, що ця основа послідовно позиціонує штучний інтелект як допоміжний інструмент, що підвищує ефективність та послідовність судової влади, водночас гарантуючи, що остаточні повноваження щодо ухвалення рішень залишаються за суддями-людьми. Водночас, аналіз визнає, що цей підхід, хоча й є узгодженим та добре відповідає сучасним технологічним реаліям, може дедалі частіше перевірятися, оскільки системи штучного інтелекту стають більш досконалими, що ставить під сумнів припущення, що є основою сучасної європейської моделі, побудованої на людському нагляді та контролі. Хоча ця система не передбачає автономних суддів-ШІ, майбутній розвиток може спонукати до повторного розгляду питання про те, чи залишаються її наявні межі адекватними для регулювання дедалі складнішої технологічної участі у судовому розгляді справ.*

Ключові слова: *суди; судовий розгляд; штучний інтелект; робот-суддя; європейське регулювання; автоматизоване ухвалення рішень; системи підтримки ухвалення рішень на основі ШІ.*

Access to Justice in Eastern Europe
ISSN 2663-0575 (Print)
ISSN 2663-0583 (Online)
http://ajee-journal.com

Research Article

EUROPEAN APPROACHES TO DIGITAL JUSTICE IN CENTRAL AND EASTERN EUROPE AND THE BALTIC STATES, WITH PERSPECTIVES FOR UKRAINE

Oleh Syniehubov, Oksana Bortnik and Olena Chernenko*

ABSTRACT

Background: *This article examines European approaches to digital justice, focusing on how supranational regulatory models—ranging from digital rights principles to operational e-justice instruments—shape national practices across Central and Eastern Europe and the Baltic States. The study conceptualises digital justice as a multidimensional phenomenon that integrates technological tools, institutional design, data governance, and human-centred values. Particular attention is given to how these European developments may inform Ukraine's justice sector reforms as the country progresses toward alignment with the EU acquis.*

Methods: *The research employs a doctrinal legal methodology combined with comparative analysis. It systematically examines EU regulatory frameworks, CEPEJ instruments, and the Digital Decade monitoring architecture alongside civil procedure legislation and e-justice platforms in eleven EU Member States. Empirical insights are drawn from the EU Justice Scoreboard 2024–2025 and Digital Decade Country*

DOI:
https://doi.org/10.33327/AJEE-18-8.S-c000162

Date of submission: 07 Nov 2025
Date of acceptance: 13 Dec 2025
Date of Publication: 30 Dec 2025

Disclaimer:

The authors declare that their opinion and views expressed in this manuscript are free of any impact of any organizations.

Reports. This methodological approach enables the identification of patterns, divergences, and implementation models, forming a basis for assessing Ukraine's digital justice trajectory.

Results and Conclusions: *The findings demonstrate that the European Union has developed a coherent, value-oriented architecture of digital justice that unites legally binding standards, interoperable technological solutions, and principles of inclusiveness, transparency, and human oversight—particularly in the context of high-risk AI. The Baltic States provide the most integrated and technologically advanced model, whereas Central and Eastern European jurisdictions exhibit more gradual or fragmented pathways. Systemic integration—characterised by mandatory electronic filing, unified data-exchange infrastructures, and machine-readable judicial data— correlates with stronger performance across EU indicators. Conversely, fragmented or parallel systems constrain accessibility, interoperability, and data-driven justice. Ukraine has achieved notable progress through UJITS, automated case management, and electronic document flow; however, substantial gaps persist in interoperability, machine-readability, user-centred design, and AI governance. Structural fragmentation and the absence of a comprehensive digital justice strategy limit its convergence with EU standards. Drawing on comparative insights, priority directions for Ukraine include full implementation of machine-readable formats (XML/JSON and ECLI), consolidation of fragmented subsystems into a unified ecosystem, mandatory digital procedures for professional participants, and the development of a rights-based AI governance framework aligned with the EU AI Act. Ensuring that digitalisation enhances–rather than restricts —access to justice requires balancing technological innovation with procedural safeguards, institutional resilience, and user inclusion.*

1 INTRODUCTION

Digitalisation and digital transformation are fundamental EU development areas that affect all spheres of public life, including justice. Within the framework of European integration and the implementation of strategic digital priorities, digital justice (e-justice) is not only a means of modernising judicial proceedings but also a means of ensuring effective access to justice, upholding human rights, and strengthening trust in legal institutions.

A clear conceptual foundation is essential for analysing the digital transformation of justice systems. In this manuscript, the terms *digitisation*, *digitalisation*, and *digital transformation* are used in accordance with contemporary scholarly systematisations. Following Derhachova and Koleshnia (2020),[1] and Saprykin (2024),[2] *digitisation* refers to the technical

1 Hanna Derhachova and Yana Koleshnia, 'Digital Business Transformation: Essence, Signs, Requirements and Technologies' (2020) 17 Economic Bulletin of National Technical University of Ukraine "Kyiv Polytechnical Institute" 283-4. doi:10.20535/2307-5651.17.2020.216367.

2 Viacheslav Saprykin, 'Digitization, Digitalization and Digital Transformation of Public Administration in Ukraine' (2024) 19(1) Bulletin of Taras Shevchenko National University of Kyiv: Public Administration 118-9. doi:10.17721/2616-9193.2024/19-19/22.

conversion of analogue information into digital formats, while *digitalisation* denotes the redesign of judicial procedures and workflows through the integration of digital technologies. In contrast, *digital transformation* is understood, in line with Elia et al. (2024)[3] and Nadkarni and Prügl (2021),[4] as a systemic and strategic reconfiguration of institutional structures, data governance models, and user experience, extending beyond technological upgrades to encompass organisational and cultural change. Within this broader framework, *e-justice* refers to the deployment of electronic tools and platforms to support judicial processes. In contrast, *digital justice* reflects a more comprehensive paradigm that integrates technological innovation with principles of fairness, transparency, accessibility, and accountability. The concepts of *justice platforms* and *data-driven justice* describe, respectively, integrated digital ecosystems that support procedural actions and interoperability, and the use of structured, machine-readable legal data to enhance decision-making, monitoring, and institutional performance. Establishing these distinctions ensures terminological precision and clarifies the analytical lens through which the subsequent comparative assessment of European jurisdictions is conducted.

EU legislation demonstrates a marked interconnection between digitalisation and the efficiency, transparency, and accessibility of justice.[5] The digitalisation of judicial processes, the development of data infrastructure, and the implementation of innovative technologies, including artificial intelligence (AI), serve as instruments to improve access to justice and strengthen the rule of law. At the same time, the EU delineates the permissible limits by classifying AI systems in judicial proceedings as high-risk. Such technologies operate exclusively under human control, in compliance with ethics, legal certainty, and accountability. The EU establishes clear boundaries for AI in judicial proceedings through risk-based classification. Under the EU AI Act, systems assisting judicial authorities in

3 Gianluca Elia and others, 'The Digital Transformation Canvas: A Conceptual Framework for Leading the Digital Transformation Process' (2024) 67(4) Business Horizons 381. doi:10.1016/j.bushor.2024.03.007.

4 Swen Nadkarni and Reinhard Prügl, 'Digital Transformation: A Review, Synthesis and Opportunities for Future Research' (2021) 71 Management Review Quarterly 233. doi:10.1007/s11301-020-00185-7.

5 European Commission, *A Digital Single Market Strategy for Europe: Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions* (COM(2015) 192 final, 6 May 2015) <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52015DC0192> accessed 25 September 2025; European Commission, *A European Strategy for Data: Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions* (COM(2020) 66 final, 19 February 2020) <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020DC0066> accessed 25 September 2025; European Commission, *2030 Digital Compass: the European way for the Digital Decade: Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions* (COM(2021) 118 final, 9 March 2021) <https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX:52021DC0118> accessed 25 September 2025; Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on Harmonised Rules on Fair Access to and Use of Data and Amending Regulation (EU) 2017/2394 and Directive (EU) 2020/1828 (Data Act) [2023] OJ L 2854/1.

interpreting facts and law or applying law to specific cases constitute high-risk AI.[6] However, Regulation (EU) 2024/1689 (Article 6(3) and Recital 61) exempts ancillary administrative functions like scheduling or transcription from this classification, as they do not materially affect legal outcomes. While UNESCO's Guidelines provide an ethical framework,[7] the EU's approach creates legally binding obligations for high-risk system providers and users. This transition from ethical guidelines to mandatory compliance ensures more effective risk management and the protection of fundamental rights in judicial AI deployment.

To identify systemic challenges and best practices for Ukraine, it is essential to analyse and characterise the experience of the EU Member States, particularly those in Central and Eastern Europe and the Baltic region, which have undergone transformations of their legal systems and recent judicial reforms. This will help to determine the ways to avoid the fragmented and inconsistent approach that is sometimes characteristic of the digital transformation of public services, and to ensure the coherent implementation of European standards, which are determinative for Ukraine, given the continued digitalisation of its judicial system.

2 METHODOLOGY

The study employs a legal methodology combined with comparative analysis to explore the normative foundations and practical implementation of European standards for digital justice, with particular emphasis on their relevance for Ukraine. The methodology integrates various analytical approaches to systematically evaluate the principles and objectives of digital transformation of justice, identifying its dynamics and structural-functional interrelations.

The study adopts a legal-analytical focus, prioritising analysis of legal acts, official monitoring instruments, and operational e-justice systems over theoretical e-justice scholarship. This methodological choice reflects the research objective: to identify normative and technical benchmarks relevant to Ukraine's legislative alignment with the EU acquis, rather than to develop conceptual frameworks of digital justice theory. EU regulatory and legal documents, as well as the national legislation of the EU Member States, serve as the primary data sources. At the same time, scientific publications, technical documentation, e-justice platforms, the EU Justice Scoreboard 2024–2025, and Digital Decade Country Reports 2024 are treated as sources that provide evidence

6 Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence and Amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) [2024] OJ L 1689/1.

7 UNESCO, *Guidelines for the Use of AI Systems in Courts and Tribunals* (UNESCO 2025). doi:10.58338/LIEY8089.

regarding the state of implementation. Scientific publications are employed to contextualise regulatory developments and provide critical interpretations where official sources require academic elaboration. For instance, analysis of European court automation trajectories (2024) informs the classification of implementation models, while CEPEJ guidelines on e-filing and digitalisation (2021) provide normative benchmarks for assessing procedural standards. This approach ensures that findings remain grounded in enforceable legal standards and replicable institutional practices, which are essential for a candidate country's accession process.

The comparative analysis of digital justice includes: current legal regulation through civil procedural legislation governing electronic filing, remote participation, and digital signatures; technical infrastructure of national platforms (E-lietas, e-Toimik, LITEKO, eŽaloby, e-Sodstvo), including e-ID authentication, registry interoperability, and machine-readability standards; and monitoring via EU Justice Scoreboard 2024-2025 and Digital Decade Country Reports indicators. Based on these criteria, national approaches are distinguished by mandatory e-filing scope (universal for all participants vs. mandatory for professionals only vs. voluntary); platform integration level (unified system vs. multiple interconnected platforms vs. fragmented portals); machine-readability standards for court decisions, including structured formats with semantic markup and ECLI identifiers; and user-centric design. This framework enables the identification of national approaches through systemic integration, gradual transformation, and fragmented adoption, and facilitates their comparative evaluation in achieving data-driven justice, accessibility, effectiveness, and digital inclusion.

The comparative study covers Central and Eastern Europe and the Baltic region: Poland, Czechia, Hungary, Slovakia, Slovenia, Croatia, Romania, Bulgaria, Estonia, Latvia, and Lithuania. Country selection is justified by four criteria: post-socialist transformation experience and European integration (EU accession 2004–2013), relevant to Ukraine's candidate country trajectory; continental legal tradition comparable to Ukraine's civil procedure framework, using which the digitalisation of justice systems is examined; diversity in digital maturity (Estonia: 95.8% digital service coverage) to fragmented implementers (Romania: 52.2%, Croatia) enabling identification of alternative development pathways and the tracking of implementation challenges; systematic data availability through EU Justice Scoreboard and Digital Decade Country Reports, ensuring methodological rigor. Primary sources were selected based on legal validity, relevance (adopted between 2020 and 2025), and applicability to justice sector regulation.

During the study, several limitations were acknowledged. Temporal constraints arise from the recent adoption of key regulatory instruments (the European e-Justice Strategy 2024-2028, the EU AI Act 2024), which limit long-term impact assessment. data availability varies across Member States, particularly regarding technical specifications of national platforms and usage statistics beyond EU Justice Scoreboard indicators. The dynamic nature of digital transformation is evident within the boundaries of the most

recent monitoring period (2024–2025). The comparative analysis is limited to civil procedure regulation and does not extend to criminal or administrative proceedings, although digitalisation patterns often converge. These limitations justify the study's focus on regulatory frameworks, official monitoring data, and operational e-justice systems rather than experimental initiatives or conceptual reforms.

The chosen methodology has enabled the formulation of well-founded conclusions: the analysis of regulatory frameworks confirms the EU's human-centric and rights-based paradigm; the comparative study identifies implementation models; and the assessment of Ukraine highlights specific directions for adapting European approaches.

The applied methodology enables a systematic examination of three core dimensions: the European regulatory framework and standards for digital justice (Section 3.1), the national practices of implementation across Central and Eastern Europe and the Baltic States (Sections 3.2–3.3), and Ukraine's challenges, gaps, and alignment prospects with EU requirements (Section 3.4). This progression from pan-European regulatory frameworks to concrete national practices and monitoring data ensures that recommendations for Ukraine are grounded in comparative evidence rather than abstract principles. The normative regulation, technical infrastructure, and practical implementation as analytical frameworks facilitate the identification of causal relationships between legal approaches, technological solutions, and measurable outcomes, as reflected in EU Justice Scoreboard indicators and Digital Decade Country Reports data.

3 RESULTS AND DISCUSSION

3.1. European Standards and Regulatory Architecture of the Digital Justice in the European Union

The approaches to digitalisation of justice in the EU are implemented directly through digital solutions within a shared politico-legal framework for EU Member States. Strategic priorities, institutional instruments, and mechanisms of the digital transformation of justice at both supranational and national levels are presented as follows.

One of the key acts that form the foundation for digital justice tools is the European Commission's 2020 Communication, *Digitalisation of Justice in the European Union: A Toolbox*.[8] Access to justice and the promotion of cooperation among Member States are central elements of the EU's area of freedom, security, and justice, as established by

8 European Commission, *Digitalisation of Justice in the European Union: Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions* (COM(2020) 710 final, 2 December 2020) <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=COM:2020:710:FIN> accessed 25 September 2025.

Article 67 of the Treaty on the Functioning of the European Union (TFEU).[9] Article 47 of the Charter of Fundamental Rights of the EU guarantees the right to an effective remedy and to a fair trial.[10] Efficient judicial systems are essential to economic growth, and digitalisation is a key element of justice modernisation. The Communication aims to support Member States in digitalising national justice systems and to enhance cross-border judicial cooperation across the EU. Its measures include online services, video conferencing, registry interconnectivity, and secure data exchange. The overarching goal is to strengthen judicial resilience through digital tools, promoting sustainable economic development, and upholding European values. The comprehensive "toolbox" of legal, financial, and IT instruments enables tailored implementation across justice systems. Digitalisation must safeguard fundamental rights, ensure accessibility for vulnerable groups, and maintain security and user trust. Ultimately, it serves to reinforce the rule of law within the European Union.[11]

The Communication identifies persistent fragmentation in the digitalisation of justice across EU Member States. Despite individual progress, disparities remain in access to electronic case files and digital evidence. Many essential services and registries are still not digitalised, with paper-based processes causing delays, higher costs, and inefficiencies. Additionally, uncoordinated national IT systems result in digital incompatibility. Voluntary initiatives like e-CODEX lacked consistency, while EU justice and security bodies faced insufficient digital infrastructure for secure evidence exchange. These shortcomings hinder cross-border cooperation and must be overcome to ensure access to justice both at the national and cross-border levels within the EU's Digital Decade framework.[12]

When discussing the challenges to the further development of digital justice, the key issues identified in the context of e-justice include technical imperfections, cybersecurity risks, dependency on digital infrastructure, limited accessibility, excessive formalisation, the erosion of the human factor, and the need for sustainable funding.[13] As Fabri (2024) notes, complex digital solutions can hinder access to justice and compromise procedural equality. The key challenge lies not in technological progress but in institutional adaptation and the need for technological simplicity.[14]

To address current challenges and enhance justice in the EU, digitalisation must follow the principles of subsidiarity and proportionality, reduce the digital divide, and respect national contexts. The EU toolkit includes funding, legislation, interoperable and accessible IT

9 Treaty on the Functioning of the European Union (consolidated version) [2012] OJ C 326/47.
10 Charter of Fundamental Rights of the European Union [2012] OJ C 326/391.
11 European Commission, *Digitalisation of Justice in the European Union* (n 8).
12 ibid
13 Vira Pyrohovska and others, 'E-Justice and the Development of Justice: Strengths, Challenges and Prospects' (2024) 16(1) Lex Humana 426.
14 Marco Fabri, 'From Court Automation to e-Justice and Beyond in Europe' (2024) 15(3) International Journal for Court Administration 7. doi:10.36745/ijca.640.

solutions, and monitoring mechanisms. Tools may be mandatory or voluntary, with a focus on accessibility for vulnerable groups, safeguarding defence rights in criminal cases, and promoting party equality in civil proceedings.[15]

As Fierbințeanu and Nemeş (2022) note, EU digital tools such as e-CODEX, the European Case Law Identifier and secure electronic communication operate not merely as technical solutions but as structural guarantees of mutual trust and procedural efficiency in cross-border cooperation. Their analysis underscores that the effectiveness of digital justice depends on the seamless interoperability of national systems and the capacity to exchange information securely, promptly, and in a user-centred manner.[16]

If the Communication *Digitalisation of justice in the European Union: A toolbox of opportunities* (2020) outlines a strategic policy approach and proposes tools for the digitalisation of justice, the Working Document provides a solid empirical foundation, a methodologically grounded study, statistical context, mapping, and a dataset on the state of digitalisation in the Member States, based on which recommendations and proposals for the digitalisation of justice were formulated.[17]

The European Declaration on Digital Rights and Principles for the Digital Decade (2023) defines a set of values-oriented benchmarks relevant to the digitalisation of justice.[18] It affirms that technological innovation must be guided by democratic principles, the rule of law, and fundamental rights. The principle of "Putting people at the centre of the digital transformation" establishes that technology in justice must enhance legal certainty and individual rights, with AI tools remaining subordinate to human control. "Solidarity and inclusion" requires that digital justice systems ensure equal access for all, particularly for disadvantaged or marginalised groups, through inclusive design and alternative formats. "Connectivity" highlights the need for high-quality internet access to support fair and effective participation in judicial processes, both nationally and across borders. "Digital education, training and skills" stresses the importance of digital literacy for users and legal professionals to ensure full and equal engagement with e-justice services. "Fair and just working conditions" addresses the impact of digitalisation on legal practitioners, calling for ethical safeguards, transparency, and protection of professional autonomy. "Digital public services online" promotes the integration of justice into broader e-government systems, supporting remote access, interoperability, and secure communication. "Interactions with algorithms and artificial intelligence systems" demand transparency, explainability, and human oversight in AI use to avoid bias and safeguard due process. "A fair digital environment" calls for transparency and openness in legal technologies, ensuring user

15 European Commission, *Digitalisation of Justice in the European Union* (n 8).

16 Gabriela Fierbințeanu and Vasile Nemeş, 'Digital Tools for Judicial Cooperation across the EU – The Benefits of Digital Technologies in Judicial Proceedings' (2022) 15(1) Challenges of the Knowledge Society 136.

17 European Commission, *Digitalisation of Justice in the European Union* (n 8).

18 European Declaration on Digital Rights and Principles for the Digital Decade [2023] OJ C 23/1.

choice and preventing monopolistic dominance in the justice sector. "Participation in the digital public space" ensures access to legal information and safeguards users from digital harms, reinforcing freedom of expression and informational inclusion. "A protected, safe and secure digital environment" emphasises cybersecurity and data protection as essential for procedural integrity and public confidence in justice. "Privacy and individual control over data" ensures that individuals retain control over their personal information, with legal protections against surveillance and unauthorised use. "Protection and empowerment of children and young people in the digital environment" affirms the need for child-sensitive justice services, focusing on clear communication, confidentiality, and digital awareness. "Sustainability" links digitalisation for justice to environmental objectives, encouraging energy-efficient infrastructure and the responsible use of digital technologies. Together, these principles articulate a rights-based, ethically grounded, and human-centred vision of digital justice that ensures both technological effectiveness and the protection of individual dignity.

An important, valid document for advancing the digital transformation of justice is the European e-Justice Strategy 2024–2028, which continues the technically oriented e-Justice strategies of previous periods (2014–2018, 2019–2023).[19] European e-Justice Strategy 2024–2028, building upon the overarching framework of the European Declaration on Digital Rights and Principles for the Digital Decade (2023), articulates a coherent set of principles guiding the digital transformation of justice in the European Union, which are categorised into substantive ("Respect for fundamental rights and principles", "Access to justice", "People centricity", "Bridging the digital divide", "Digital empowerment of users", "Sustainability") and operational ("Once-only principle", "Digital by default", "Interoperability and cybersecurity", "Dynamic justice", "Data-driven justice", "Open-source") and jointly promote a cohesive, rights-based, and future-oriented approach of digital justice, rooted in legality, inclusiveness, and sustainability.[20] In addition, the act outlines key challenges and sets strategic and operational goals for digital justice, supported by an action plan and monitoring mechanisms at both EU and national levels. These goals include expanding access through universal digital services and bridging the digital divide; enhancing cross-border cooperation via interoperability and real-time services; improving efficiency through data-driven solutions and digitalisation of in-person processes; and fostering innovation by implementing advanced technologies and sharing best practices.[21]

As Makauskaite-Samuolė (2025) observes, the EU AI Act forms part of a broader transparency architecture aimed at preventing opaque algorithmic practices in judicial systems. Its requirements on documentation, traceability and explicability function as structural safeguards of judicial accountability, reinforcing the principle that AI must

19 Council of the European Union, *Draft Strategy on European e-Justice 2014-2018* [2013] OJ C 376/6; Council of the European Union, *2019-2023 Strategy on e-Justice* [2019] OJ C 96/3.

20 Council of the European Union, *European e-Justice Strategy 2024–2028* [2025] OJ C 437/1.

21 ibid

remain fully subordinate to human oversight and embedded within rights-protecting institutional frameworks.[22]

The digitalisation of justice is also specifically addressed by the European Commission for the Efficiency of Justice (CEPEJ). In its Action Plan "Digitalisation for Better Justice" 2022–2025, CEPEJ identifies as a priority the support of states and courts in successfully transitioning to the digitalisation of justice in accordance with European standards.[23] CEPEJ promotes humane, effective, and high-quality justice through the following principles: efficiency (digitalisation of judicial management), transparency (greater public awareness of justice), shared justice (digital interconnection of participants), humane and human-centred justice (support and training for legal professionals), informed justice (use of evaluation results), and responsible and responsive CEPEJ (accessible tools and demonstrated expertise).[24]

Regulation (EU) 2022/850 positions e-CODEX as a central instrument for cross-border judicial cooperation in civil and criminal matters. This decentralised system enables the secure, scalable, and interoperable exchange of electronic documents, evidence, and metadata between national judicial authorities. Aimed at enhancing access to justice for individuals and businesses, e-CODEX operates in accordance with the highest EU standards on cybersecurity and data protection, including the GDPR[25] and eIDAS[26]. Although its use in civil matters remains voluntary, participating Member States must ensure technical compliance and regularly monitor system performance through non-personal data reporting, thereby reinforcing efficiency and institutional transparency.[27]

22 Gintare Makauskaite-Samuole, 'Transparency in the Labyrinths of the EU AI Act: Smart or Disbalanced?' (2025) 8(2) Access to Justice in Eastern Europe 38. doi:10.33327/ajee-18-8.2-a000105.

23 CEPEJ, *2022–2025 CEPEJ Action Plan: "Digitalisation for a Better Justice"* (CEPEJ(2021)12Final, 9 December 2021 <https://rm.coe.int/cepej-2021-12-en-cepej-action-plan-2022-2025-digitalisation-justice/1680a4cf2c> accessed 25 September 2025; Council of Europe, *European Convention on Human Rights: as amended by Protocols Nos 11, 14 and 15, supplemented by Protocols Nos 1, 4, 6, 7, 12, 13 and 16* (ECHR 2013).

24 CEPEJ, *2022–2025 CEPEJ Action Plan* (n 23).

25 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation - GDPR) [2016] OJ L 119/1.

26 Regulation (EU) 910/2014 of the European Parliament and of the Council of 23 July 2014 on Electronic Identification and Trust Services for Electronic Transactions in the Internal Market [2014] OJ L 257/73; Regulation (EU) 2024/1183 of the European Parliament and of the Council of 11 April 2024 Amending Regulation (EU) No 910/2014 as Regards Establishing the European Digital Identity Framework [2024] OJ L 1183/1.

27 Regulation (EU) 2022/850 of the European Parliament and of the Council of 30 May 2022 on a Computerised System for the Cross-Border Electronic Exchange of Data in the Area of Judicial Cooperation in Civil and Criminal Matters (e-CODEX system), and Amending Regulation (EU) 2018/1726 [2022] OJ L 150/1.

Regulation (EU) 2023/2844 provides a unified legal framework for the digitalisation of judicial cooperation procedures in civil, commercial, and criminal matters with a cross-border element. It harmonises the use of electronic communication between judicial authorities and parties, establishes legal standards for videoconferencing, electronic signatures and seals, and defines the legal effects of electronic documents. The Regulation also introduces mechanisms for electronic payment of court fees. Its implementation is phased, beginning in 2025 with provisions on remote communication, and extending further in 2028 and 2031.[28] As noted in legal scholarship, the Regulation's reference to "access to justice" reflects its broader scope, which extends beyond the digitalisation of cross-border judicial cooperation. It also covers procedural mechanisms in European civil and criminal justice designed to facilitate access to justice in transnational disputes, even without the direct involvement of international cooperation mechanisms.[29]

It should also be noted that in 2013, the EU Justice Scoreboard was initiated as a tool for comparative analysis of the efficiency, quality, and independence of judicial systems, with a particular focus on recovery following the 2008 financial crisis.[30] The EU Justice Scoreboard serves as an instrument to support the rule of law within the EU, providing a comparative basis for assessing judicial reforms, enhancing institutional transparency, integrating market indicators into justice analysis, and fostering coordination within the European Semester and the Rule of Law Mechanism.[31]

The 2025 edition of the EU Justice Scoreboard functions as a comparative monitoring tool for Member States' judicial systems, structured around three core dimensions: efficiency, quality, and independence. Efficiency is assessed through indicators such as caseloads, disposition time, and clearance rates. Quality is measured by the digitalisation of judicial services, accessibility for citizens and professionals, and support for vulnerable groups. Independence is evaluated through public perception and the governance of judicial appointments, while new indicators in 2025 extend to oversight bodies and market-related

28 Regulation (EU) 2023/2844 of the European Parliament and of the Council of 13 December 2023 on the Digitalisation of Judicial Cooperation and Access to Justice in Cross-Border Civil, Commercial and Criminal Matters, and Amending Certain Acts in the Field of Judicial Cooperation [2023] OJ L 2844/1.

29 Fernando Gascón Inchausti, 'The New Regulation on the Digitalisation of Judicial Cooperation in the European Union: Something Old, Something New, Something Borrowed and Something Blue' (2023) 24 ERA Forum 535. doi:10.1007/s12027-024-00782-z.

30 European Commission, 'EU Justice Scoreboard: European Commission Broadens the Scope of its Analysis of Member States' justice systems: Press release' (*European Commission*, 27 March 2013) <https://ec.europa.eu/commission/presscorner/detail/en/ip_13_285> accessed 25 September 2025.

31 'EU Justice Scoreboard' (*European Commission*, July 2025) <https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/upholding-rule-law/eu-justice-scoreboard_en> accessed 25 September 2025; Jenny Gesley, 'FALQs: The Rule of Law in the European Union' (*Library of Congress blogs*, 12 August 2022) <https://blogs.loc.gov/law/2022/08/falqs-the-rule-of-law-in-the-european-union/> accessed 25 September 2025.

institutions. A distinct focus in the 2025 EU Justice Scoreboard is placed on the digitalisation of justice, which is viewed as a key driver of institutional modernisation. The data indicate a rise in Member States allowing digital evidence (from 6 to 9) and enabling fully online initiation of judicial proceedings (in 26 countries). The EU Justice Scoreboard 2025 confirms steady progress in justice digitalisation, improved service quality, strengthened judicial independence, and greater legal harmonisation across the EU. Its findings provide an essential evidence base for assessing legal reforms and guiding strategic planning within the EU's rule of law framework.[32]

The analysis of EU legal acts on the digitalisation of justice demonstrates a consistent development of a unified regulatory and instrumental framework for the digital transformation of national judicial systems within the EU's legal space. The key dimensions of this process include: value-based orientation (the European Declaration on Digital Rights and Principles for the Digital Decade (2023), strategic planning (the European e-Justice Strategy 2024–2028), operationalisation of digital tools (the 2020 Communication "Digitalisation of Justice in the European Union: A Toolbox" and e-CODEX), methodological support (the CEPEJ Action Plan), and progress monitoring (the EU Justice Scoreboard). Digitalisation is recognised as an instrument of modernisation aimed at ensuring equal access to justice, enhancing the efficiency of cross-border cooperation, strengthening the resilience of judicial systems, and upholding fundamental rights. At the same time, emphasis is placed on bridging the digital divide, ensuring the interoperability of technical solutions, securing data exchange, and preserving the human dimension of justice. The digital transformation of the judiciary in the EU is multidimensional, encompassing legal, organisational, technological, and social aspects, and is aimed at the long-term consolidation of the rule of law within Europe's Digital Decade. Similarly, Razmetaeva and Razmetaev (2021) contend that digital justice must be assessed beyond technological efficiency, with attention to risks such as expanded surveillance, reduced procedural fairness, and weakened individual autonomy. They emphasise that digitalisation can enhance rights only where institutional safeguards preserve the human-centred nature of adjudication.[33]

32 European Commission, *The 2025 EU Justice Scoreboard: Communication from the Commission to the European Parliament, the Council, the European Central Bank, the European Economic and Social Committee and the Committee of the Regions* (COM (2025) 375, Publications Office of the EU 2025).

33 Yulia Razmetaeva and Sergiy Razmetaev, 'Justice in the Digital Age: Technological Solutions, Hidden Threats and Enticing Opportunities' (2021) 4(2) Access to Justice in Eastern Europe 104. doi:10.33327/AJEE-18-4.2-a000061.

3.2. National Practices of Digital Justice in the Countries of Central and Eastern Europe

The effectiveness of European standards depends on their practical implementation at the national level. The realisation of the European paradigm of digital justice demonstrates significant variability in approaches among the EU Member States. The study, alongside digital platforms that mediate the possibilities of judicial proceedings and the implementation of the principles and objectives of digital justice, encompasses civil litigation due to the following factors: the significant share of judicial workload and the involvement of the broadest range of individuals, the fullest manifestation of the challenges of ensuring equal access to justice, the greatest diversity of technological solutions, and the variety of approaches to safeguarding procedural guarantees among the EU Member States. For comparative analysis, the Member States of Central and Eastern Europe and the Baltic region have been selected, as they have undergone transitions to democratic legal states and EU membership over recent decades. Their experience in implementing European principles of digital justice within the context of relatively recent judicial modernisation offers an opportunity to assess the challenges of adapting European standards during European integration, providing valuable insights for Ukraine as a candidate country.

Comparative scholarship likewise highlights the regional relevance of this analysis. Piatyhora (2024) observes that, although EU-aligned standards offer a coherent normative model, e-justice implementation in Central and Eastern Europe remains uneven due to infrastructural, organisational, and regulatory fragmentation.[34] This underscores the value of examining Central and Eastern Europe jurisdictions as a testing ground for post-transition digital justice reform.

Poland. The 2024 EU Justice Scoreboard indicates that Poland meets 8 of 12 indicators for online judicial information, though chatbots, online legal aid, citizens' rights resources, and online training are lacking. The legal framework for digital technologies is moderately developed, but blockchain and AI are not applied. Judicial decisions lack machine-readable modelling, metadata, ECLI support, and automated downloads. Digital support for initiating and tracking civil cases remains weaker than in administrative proceedings.[35] As of 2025, the majority of these indicators remain unchanged.[36]

34 Kristina V Piatyhora, 'International Experience of Implementing E-Justice: Best Practices and Challenges' (2024) 1(25) Theory and Practice of Jurisprudence 107. doi:10.21564/2225-6555.2024.1(25).305796; Kristina V Piatyhora, 'E-justice in Administrative Process: European Standards and Foreign Experience' (2024) 2(24) Theory and Practice of Jurisprudence 39. doi:10.21564/2225-6555.2023.2.293064.

35 European Commission, *The 2024 EU Justice Scoreboard: Communication from the Commission to the European Parliament, the Council, the European Central Bank, the European Economic and Social Committee and the Committee of the Regions* (COM (2024) 388, Publications Office of the EU 2024) 35-7, 40.

36 European Commission, *The 2025 EU Justice Scoreboard* (n 32) 33-40.

Based on the Digital Decade Country Report 2024, Poland's digitalisation indicators for public services are at a moderate level, scoring 63.7 for citizens and 72.9 for businesses, compared to the EU averages of 79.4 and 85.4 respectively, with only moderate growth observed.[37]

According to the Polish Code of Civil Procedure,[38] electronic access to case materials is available, including the ability to review documents, obtain copies, duplicates, or extracts (Article 9), and to track the progress of case proceedings. Interoperability, along with inclusion, empowerment, and adherence to the "once-only" principle, is facilitated by the rule that professional representatives, authorities, and individuals specified in Article 67 are not required to confirm their authorisations if this information is already available to the court through official registers (Article 68). In cases defined by law or where chosen by a party, procedural documents must be submitted exclusively via the Electronic Information System (EIS) (Information Portal[39]) (Article 125), with electronically certified copies of attachments (Article 128). Court hearings may be conducted remotely, either at the initiative of the presiding judge or upon request by a participant (submitted within seven days of the summons' delivery date). Persons deprived of liberty are obliged to participate remotely. The judge may require physical presence in another court building to ensure the necessary technical conditions (Article 151). Within EIS proceedings, judgments are recorded using the judge's QES (§ 4, Article 324).[40]

According to Wrzaszcz (2023), Poland's e-justice trajectory reveals a tension between formal regulatory preparedness and practical implementation. Despite detailed rules for electronic communication, progress is constrained by uneven institutional capacity, limited user readiness, and the lack of a unified digital justice ecosystem.[41]

Electronic delivery of documents is carried out in accordance with the Polish Code of Civil Procedure and the special act issued by the Minister of Justice.[42] Starting on 1 January 2025, electronic communication with a public authority, by default, will take

37 European Commission, *Digital Decade Country Report 2024: Poland* (Publications Office of the EU 2024) <https://ec.europa.eu/newsroom/dae/redirection/document/106821> accessed 28 September 2025.

38 Code of Civil Procedure of the Republic of Poland of 17 November 1964 'Kodeks postępowania cywilnego' (amended 1 January 2025) <https://sip.lex.pl/akty-prawne/dzu-dziennik-ustaw/kodeks-postepowania-cywilnego-16786199> accessed 28 September 2025.

39 Ministerstwo Sprawiedliwości, *Podręcznik Użytkownika: Portalu Informacyjnego Sądów Powszechnych 2.5* (*Portal Informacyjny Sądów Powszechnych*, 17 June 2025) <https://portal.wroclaw.sa.gov.pl> accessed 28 September 2025.

40 Code of Civil Procedure of the Republic of Poland (n 38).

41 Paweł Wrzaszcz, 'E-justice in Poland – Polish Experiences' (2023) 16(1) Teka Komisji Prawniczej PAN Oddział w Lublinie 381. doi:10.32084/tkp.5288.

42 Regulation of the Minister of Justice of the Republic of Poland 'On the Procedure and Method of Electronic Service' (20 October 2015) <https://sip.lex.pl/akty-prawne/dzu-dziennik-ustaw/tryb-i-sposob-dokonywania-doreczen-elektronicznych-18235252> accessed 28 September 2025.

place via the state-registered electronic delivery service.[43] This method constitutes the general mechanism for electronic delivery, either at the recipient's choice or by initiating proceedings in electronic form (Article 131-1).[44]

For electronic simplified proceedings, under the pan-European procedure,[45] the E-Court System (EPU) for simplified proceedings (Articles 505.28–505.39) provides for the creation of a user account via a registration form, accompanied by detailed instructions for various categories of users: computer setup, registration, login, profiles for judicial staff, bailiffs, professional lawyers, basic users, and mass claimants. The guides embody the principles of inclusion, accessibility, and security by explaining authentication procedures. Claims are signed either with a QES or a simple electronic signature, allowing automatic, free signing with an active account. The Minister of Justice determines the formats of documents, taking into account the minimum requirements for public registers (§ 5, Article 129). Within the EPU, documents submitted by any other means have no legal effect, which is consistently indicated on the website.[46]

Romania. The 2024 EU Justice Scoreboard notes Romania's progress in online information access and case management but highlights persistent gaps, including incomplete integration of judicial actors, limited digital access for defendants, and insufficient use of advanced technologies. To align with EU standards, Romania should adopt chatbots for information services, apply AI in judicial processes, implement machine-readable judgment models, include metadata in decisions, and enable automatic publication of judicial information.[47] Procedural rules defining the use of digital technologies in courts are at the stage of development and improvement, particularly regarding the electronic submission of documents and confirmation of receipt. The 2025 EU Justice Scoreboard predicts progress in ICT regulation; however, chatbots, online legal aid assessment, case outcome evaluation tools, and interactive learning platforms are still absent.[48]

Indicators of digitalisation of public services in Romania remain low (52.2 for citizens and 50 for businesses, compared to 79.4 and 85.4, respectively, in the EU); however, there is a positive upward trend. Romanian citizens attach significant importance to digitalisation for accessing public services (64%). In 2023, only 1.6% of Romanians used the ROeID electronic identification scheme with a high level of assurance to access public services, compared to 35.8% across the EU. In January 2024, a law was adopted obliging public authorities to

43 Rafał Bułach, 'E-doręczenia w administracji publicznej' (*Wolters Kluwer*, 13 August 2024) <https://www.wolterskluwer.com/pl-pl/expert-insights/e-doreczenia-w-administracji-publicznej> accessed 28 September 2025.

44 Code of Civil Procedure of the Republic of Poland (n 38).

45 *Elektroniczne Postępowanie Upominawcze (Rzeczpospolita Polska)* <https://www.e-sad.gov.pl> accessed 28 September 2025.

46 Code of Civil Procedure of the Republic of Poland (n 38).

47 European Commission, *The 2024 EU Justice Scoreboard* (n 35) 40.

48 European Commission, *The 2025 EU Justice Scoreboard* (n 32) 33-5.

publish information and application forms in electronic format, to accept electronic copies of identity documents, and to refrain from requiring paper-based documentation.[49]

The digitalisation of civil proceedings in Romania includes electronic communication as the preferred method of delivery, while maintaining traditional methods (paper, fax) if such details have been provided by a party (Article 154 of the Civil Procedure Code of Romana)[50] Electronic notifications are accompanied by a QES or an advanced electronic signature, which substitutes the court's stamp (Article 154(6)). Decisions are delivered by email, signed electronically, provided the party's electronic contact details are available (Article 154-1(1)). Documents are deemed received when the system notifies the sender of the recipient's receipt (Articles 154(6-1) and 154-1). Where electronic delivery is not possible, alternative methods are applied (Article 154-1(3)). Courts have direct access to electronic databases of public authorities to retrieve necessary data (Article 154(8)). Digital technologies accelerate court proceedings and enable prompt summonses (Article 241). Electronic documents are admissible provided they meet the same requirements as written documents and comply with the provisions of a special law that also regulates the application of electronic signatures (Articles 266–268).

In Romania, the National Electronic File has been introduced.[51] This IT platform, provided by the Ministry of Justice, is designed to facilitate online access for parties and participants, or, where applicable, their legal representatives, to procedural documents and case-related information pending before the courts.[52] For authentication and access to judicial data—such as requests, summonses, and court documents—users must provide an email address and mobile phone number. Requests are completed and submitted online through a personal user account. Additionally, separate portals for individual appellate courts and tribunals—such as the Craiova Court of Appeal[53] and the Bucharest Tribunal[54] – have implemented their own "Electronic File" systems. The portal of the High Court of Cassation and Justice of Romania[55] also provides access to case files. The National Electronic File guidelines govern user registration, interface use, password recovery, data protection, and

49 European Commission, *Digital Decade Country Report 2024: Romania* (Publications Office of the EU 2024) 4-5, 22 <https://ec.europa.eu/newsroom/dae/redirection/document/106692> accessed 28 September 2025.

50 Civil Procedure Code of Romania No 134/2010 of 1 July 2010 'Codul De Procedură Civilă' (amended 10 April 2015) <https://legislatie.just.ro/public/detaliidocument/140271> accessed 28 September 2025.

51 *Dosarul Electronic Național (DEN): Serviciul electronic de acces la justiția României* <https://den.just.ro> accessed 28 September 2025.

52 'Informații utile' (*Dosarul Electronic Național (DEN)*, 2025) <https://den.just.ro/informatii> accessed 28 September 2025.

53 'Dosar Electronic' (*Curtea de Apel Craiova*, 2025) <https://www.curteadeapelcraiova.eu/dosar-electronic> accessed 28 September 2025.

54 'Dosar Electronic' (*Tribunalul București*, 2025) <https://tribunalulbucuresti.ro/index.php/informatii/anunturi/anunt-dosar-electron> accessed 28 September 2025.

55 'Dosar Electronic' (*Înalta Curte de Casație și Justiție*, 2025) <https://www.iccj.ro/acasa/dosar-electronic/> accessed 28 September 2025.

cookie policies. Access, however, requires an active court case, registration data, and two-factor authentication via a mobile app. Transitional challenges persist, including parallel use of outdated applications, restricted online access to archived files, and difficulties for legal entities with shared contact data.

Czechia. The 2024 EU Justice Scoreboard shows that Czechia's judicial digitalisation remains moderate, with electronic case management still developing. Digital limitations persist in criminal and civil proceedings, particularly regarding procedural actions, fee payments, and access to closed cases. Blockchain and AI are not applied, court decisions lack machine-readable formats, and anonymisation is only partially automated. Although all decisions are available online, they remain unsuitable for automated data processing.[56] According to 2025 data, progress has been made in anonymisation and pseudonymisation, as well as in tracking the progress of criminal cases. However, court decisions still fail to comply with machine-readable modelling standards.[57]

According to the Digital Decade Country Report 2024, significant efforts are being made in Czechia to digitalise key public services. A total of 78% of the Czech population believes that the digitalisation of everyday public and private services facilitates their lives. The indicators of digitalisation of public services for citizens (76.3) and businesses (83.8) in Czechia lag behind the EU averages (79.4 and 85.4, respectively). There is widespread use of e-ID, with ten banking institutions providing related services. To enhance the security of IT tools in public administration, information systems considered critical were updated in 2023, and solutions for threat detection were implemented (for example, within the police system). Czechia has been advised to provide more information regarding the implementation of digital rights and principles.[58]

Digital facilities for filing applications and accessing case information are provided via email, the data mailbox, the ePodatelna web application, or an electronic form. The website of the Ministry of Justice of Czechia contains a link to the Citizen Identification Portal[59], where a Czech data mailbox can be set up.[60] This mailbox serves as a key tool for electronic interaction with public authorities in Czechia. The Information System of Data Mailboxes (ISDS) has been operational since 1 July 2009. The Czech Code of Civil Procedure provides for delivery via the Datová schránka (§ 45–47, § 49): if a document is not delivered during the hearing, the court sends it through the data mailbox, which is

56 European Commission, *The 2024 EU Justice Scoreboard* (n 35) 36-40.

57 European Commission, *The 2025 EU Justice Scoreboard* (n 32) 38-40.

58 European Commission, *Digital Decade Country Report 2024: Czechia* (Publications Office of the EU 2024) 3, 5, 17, 23 <https://ec.europa.eu/newsroom/dae/redirection/document/106708> accessed 28 September 2025.

59 *Portál Identity občana (Česká Republika)* <https://www.identita.gov.cz/> accessed 28 September 2025.

60 *Datová schránka (Česká Republika)* <https://info.mojedatovaschranka.cz/info/cs/> accessed 28 September 2025.

considered equivalent to personal delivery. When using email, a signed confirmation from the recipient is required (§ 50f(2)). If delivery via the data mailbox is not possible, email may be used with the recipient's prior consent, provided receipt is confirmed within three days; otherwise, the delivery is invalid (§ 47(2–3)). In urgent cases, telephone or fax may be used (§ 51(1)). Where a party is represented, documents are sent to the representative, including electronically (§ 50b(2))[61]. The ePodatelna web application, operational since 2007, enables the submission of electronic applications and attachments, provided they bear a QES. The document web wizard allows users to fill in an electronic form and generate a PDF file.[62] Exclusively in electronic form and using established templates, it is possible to submit proposals for the issuance of an electronic payment order (§ 174a–175)[63], certain documents required for initiating insolvency proceedings, requests for specific information to public registers, and particular documents intended for catchpoles and notaries.[64] Instructions for the relevant public templates are provided to facilitate their use.

Croatia. The 2024 EU Justice Scoreboard indicates a low level of digital implementation, with limited secure communication between courts and detention facilities, no use of blockchain, and no confidential remote contact between suspects and attorneys. Electronic document service, online legal aid assessment, and remote court participation are unavailable. Automated uploading with metadata, algorithmic anonymisation, and AI tools are absent, while the machine-readability of decisions remains below the EU average[65], although seminars on automatic anonymisation are being conducted to support default digitalisation.[66] The 2025 EU Justice Scoreboard reports progress in civil, commercial, and particularly administrative proceedings, marked by the introduction of online legal aid, digital evidence formats, fully remote hearings, and the deployment of chatbots.[67]

The submission of claims in civil proceedings is carried out via the "e-Komunikacija" system (Article 106a of the Civil Procedure Act of Croatia),[68] integrated into the e-Građanin

61 Civil Procedure Code of the Czech Republic No 99/1963 Coll of 4 December 1963 'Občanský soudní řád' (amended 1 July 2025) <https://www.zakonyprolidi.cz/cs/1963-99> 28 September 2025.

62 'Informace – příjem podání v elektronické podobě soudy' (*Ministerstvo spravedlnosti České republiky*, 2025) <https://epodatelna.justice.cz/ePodatelna/info-el-podani> accessed 28 September 2025.

63 Civil Procedure Code of the Czech Republic (n 61).

64 'Formuláře ke stažení' (*Ministerstvo spravedlnosti České republiky*, 2025) <https://epodatelna.justice.cz/ePodatelna/form> accessed 28 September 2025.

65 European Commission, *The 2024 EU Justice Scoreboard* (n 35) 37-40.

66 'Croatia to use AI for Greater Transparency and Efficiency in Judiciary' (*Council of Europe: Implementation of standards - Human Rights, Justice and Legal Co-Operation*, 6 March 2024) <https://www.coe.int/en/web/implementation/-/croatia-to-use-ai-for-greater-transparency-and-efficiency-in-judiciary> accessed 28 September 2025.

67 European Commission, *The 2025 EU Justice Scoreboard* (n 32) 33-4, 37.

68 Civil Procedure Act of the Republic of Croatia 'Zakon o parničnom postupku' (amended 30 December 2023) <https://www.zakon.hr/z/134/zakon-o-parni%C4%8Dnom-postupku> accessed 28 September 2025.

portal,[69] provided they are signed with a QES in accordance with the special law. Access to the system requires identification through the National Identification System. State authorities, the prosecution service, attorneys, notaries, court experts, and other legal professionals are obliged to submit documents electronically. In cases where a submission is inadmissible, the applicant is notified of the possibility of making corrections. Electronic delivery is carried out via the portal, either with the party's consent or automatically upon electronic submission of a claim (Article 133d). Documents are sent to a secure inbox with a 15-day confirmation period, after which delivery is deemed completed. Both electronic and scanned documents bearing a qualified seal may be delivered (Article 143). If electronic delivery is not possible, written form is used instead. Forms for claims under the European Small Claims Procedure may be submitted either by fax or electronically (Article 507).[70]

The "e-Komunikacija" system serves judicial professionals and private users but faces significant access challenges. These include mandatory use of QES, strict security standards, automatic deletion of unverified addresses, and user liability for technical failures. The system remains fragmented: land cases use a separate ZIS/OSS platform, and major municipal courts are unconnected. Managed via eSpis, from which data are transferred, errors require direct court contact, and unavailable eSpis decisions cannot be accessed. Additional complexity arises from the separate "e-Ovlaštenja" system for employee authorisation. Only QES-signed PDF documents up to 20 MB are accepted; scanned signatures are rejected.[71]

Bulgaria. The 2024 EU Justice Scoreboard shows that Bulgaria offers chatbots for judicial information but lacks online case assessment and legal aid eligibility checks. While most indicators exceed the EU average, electronic communication and advanced digital technologies remain poorly implemented. In criminal proceedings, victims and defendants can access case files and apply for legal aid online. Bulgaria nonetheless achieves the highest machine-readability of court decisions in the EU.[72] The 2025 EU Justice Scoreboard notes major progress in civil proceedings, especially in electronic case initiation and tracking, while other indicators show little change.[73]

According to the Digital Decade Country Report 2024, Bulgaria recorded an increase in the digitalisation of public services for citizens (13.4% between 2022 and 2023). However, the absolute value (67.5) remains below the EU average (79.4), while the score for businesses is 91.9, which is above the EU average (85.4). Legislative amendments in 2023 introduced the "Once-Only Principle". At the time of the report, the National Electronic

69 e-Komunikacija: Elektroničke komunikacije sudionika sudskih postupaka sa sudovima' (*e-Građani: Informacije i usluge*, 2025) <https://e-komunikacija.pravosudje.hr/> accessed 28 September 2025.

70 Civil Procedure Act of the Republic of Croatia (n 68).

71 Ministarstvo Pravosuđa i Uprave Republike Hrvatske, *E-Komunikacija: Priručnik za korištenje* (verzija 0.9, listopad 2025) 74-5, 83 <https://e-komunikacija.pravosudje.hr/> accessed 28 September 2025.

72 European Commission, *The 2024 EU Justice Scoreboard* (n 35) 35-6, 38, 40.

73 European Commission, *The 2025 EU Justice Scoreboard* (n 32) 37.

Identification System was still under development. In 2023, 6.09% of Bulgaria's population used existing e-ID solutions, while 35.5% of the population possessed basic digital skills (compared to 55.6% in the EU).[74] There is an underdeveloped culture of data sharing within public administrations, high cybersecurity risks, and a low proportion of public services provided entirely online.

Starting on 1 July 2025, the judicial process in Bulgaria should undergo significant changes: more than half of civil cases in local courts will transition to a fully electronic format with centralised allocation, allowing cases to be adjudicated by a judge in any region, regardless of traditional rules of territorial jurisdiction.[75] These changes affect all participants in judicial proceedings, including legal entities, individuals, attorneys, and other participants with professional roles. Borgesano et al. (2025), in their systematic review of Justice 5.0 frameworks, argue that large-scale transitions towards fully digital judicial ecosystems require robust governance mechanisms to prevent institutional overload and ensure users' ability to adapt.[76] Bulgaria's ambitious reforms must therefore be evaluated not only through the lens of efficiency but also in light of their capacity to sustain human-centred adjudication and avoid creating new forms of procedural vulnerability.

The Bulgarian Code of Civil Procedure provides for electronic delivery through the Unified Portal of Electronic Justice, qualified electronic mail in accordance with Regulation (EU) No 910/2014, or a secure delivery system, with the possibility of withdrawing consent and reverting to traditional methods if electronic service is not feasible (Art. 38). If an address is changed without prior notification, documents are considered delivered once attached to the case file, and confirmation is recorded in the system or provided by a qualified service provider (Art. 44). Procedural actions are carried out electronically in accordance with Regulation (EU) No 910/2014 and require a qualified electronic signature, with automatic verification of standards, formats, and identification (Arts. 102g, 183). Information on electronic functionalities is provided on court websites and the justice portal (Art. 102v). Electronic payment of court fees is enabled, with a 15% discount applied (Arts. 71, 73), while submission of certified paper copies is also permitted (Art. 184). Since 1 July 2025, writ proceedings should be conducted electronically, unless otherwise provided (Article 409a). The writ contains the judge's electronic signature (Article 412), and enforcement officers gain access through the

74 European Commission, *Digital Decade Country Report 2024: Bulgaria* (Publications Office of the EU 2024) 4, 28-30 <https://ec.europa.eu/newsroom/dae/redirection/document/106715> accessed 2 October 2025.

75 'Information Service: User Guide' (*Electronic Court Cases: A Single Portal for Electronic Justice in the Republic of Bulgaria*, 2023) <https://ecase.justice.bg/help/epep-manual-public.pdf> accessed 2 October 2025.

76 Francesco Borgesano and others, 'Artificial intelligence and justice: a systematic literature review and future research perspectives on Justice 5.0' (2025) 28(11) European Journal of Innovation Management 349. doi:10.1108/EJIM-01-2025-0117.

Unified Portal (Article 418). Public auctions in enforcement proceedings are also conducted via the Ministry of Justice's online platform (Article 501a).[77]

Hungary. The 2024 EU Justice Scoreboard indicates that Hungary performs well in judicial digitalisation overall. However, there is no online data on interactive training, legal aid simulations, or chatbots. Judicial decisions remain non-machine-readable, lacking automatic downloads, metadata integration, algorithmic anonymisation, and standardised modelling.[78] The 2025 EU Justice Scoreboard reports on the recognition of the admissibility of fully digital evidence, with other indicators remaining unchanged.[79]

Hungary demonstrates significant progress in the digitalisation of public services. Over 70% of the Hungarian population possesses a national e-ID, although its use remains largely limited within the country. The Digital Citizenship Mobile App is intended to enable cross-border use of digital identification and related services. However, a key limiting factor for unlocking the full digital potential of public services, including judicial processes, lies in the existing legal framework. Current regulations require personal presence for certain procedural actions, which restricts the seamless implementation of fully digital solutions.[80]

The Hungarian Act of Civil Procedure provides detailed regulation of electronic communication, the submission of documents, delivery, and remote participation in court hearings. Electronic service constitutes the primary method for parties engaged in electronic communication, with mandatory electronic filing required for professional representatives (§ 608(1)). For other participants, electronic communication remains voluntary at any stage of the proceedings (§ 605(1)). Documents are digitised and retained in paper form (§ 613(1)). A "fiction of delivery" is established in cases of refusal to accept delivery or failed delivery attempts, with explicit legal consequences for non-compliance with electronic formats (§§ 137, 618). The evidential value of both public and private electronic documents bearing an electronic signature is formally recognised (§ 323(3a)). Electronic means are also employed for court hearings, witness examinations, and the inspection of evidence, with participants permitted to use their own devices for these purposes (§§ 622–627)[81].

Slovakia. The 2024 EU Justice Scoreboard shows that Slovakia achieves high levels of judicial digitalisation. It provides comprehensive online access to information and procedural ICT rules across all proceedings, with strong electronic communication within

77 Civil Procedure Code of the Republic of Bulgaria No 59/2007 (effective 1 March 2008, amended 8 July 2025) <https://lex.bg/laws/ldoc/2135558368> accessed 2 October 2025.

78 European Commission, *The 2024 EU Justice Scoreboard* (n 35) 35, 40.

79 European Commission, *The 2025 EU Justice Scoreboard* (n 32) 33-40.

80 European Commission, *Digital Decade Country Report 2024: Hungary* (Publications Office of the EU 2024) 6, 21-2 <https://ec.europa.eu/newsroom/dae/redirection/document/106697> accessed 2 October 2025.

81 Civil Procedure Code of Hungary No CXXX/2016 'A polgári perrendtartásról' (amended 19 August 2025) <https://net.jogtar.hu/jogszabaly?docid=A1600130.TV> accessed 2 October 2025.

courts. Advanced digital tools for case initiation, management, and support exceed EU benchmarks, and machine-readability standards are fully met.[82] In 2025, Slovakia improved access to closed civil cases and introduced fully remote hearings in criminal proceedings.[83]

Electronic filing of claims is facilitated via the Slovak Ministry of Justice portal[84] or directly through the designated platform.[85] Particular emphasis is placed on inclusivity and the accessibility of information regarding electronic public services. Thus, the authorisation process and related procedures are designed to ensure user-friendly access, offering the option to engage with information through video tutorials[86] and PDF-based instructions[87] to accommodate diverse user needs.

The Slovak Code of Civil Procedure stipulates that unauthorised electronic submissions must be additionally filed in paper form or in an authorised electronic form within ten days (§ 125(2)). Priority is given to electronic delivery via an authorised electronic mailbox or to an electronic address; this service is deemed completed after three days, even if the document has not been read (§ 105). When a decision is made without a hearing, the court informs the parties of the place and time of its publication on the official notice board and the court's website (§ 219(3))[88]. Simplified proceedings for monetary disputes are conducted exclusively in electronic form before the District Court of Banská Bystrica.[89]

Assessment of the "eŽaloby" system against the 2023 Declaration on Digital Rights and Principles highlights its accessibility, inclusiveness, and transparency, allowing users to track submissions and receive status updates via the slovensko.sk portal and email. Privacy and data protection comply with Act No. 122/2013 through QES and secure eID mailboxes. Security and reliability are ensured by qualified signatures, document

82 European Commission, *The 2024 EU Justice Scoreboard* (n 35) 35-8, 40.

83 European Commission, *The 2025 EU Justice Scoreboard* (n 32) 34, 37.

84 'Ežaloby' (*Ministerstvo Spravodlivosti Slovenskej republiky*, 2025) <https://obcan.justice.sk/ezaloby> accessed 2 October 2025.

85 'Elektronické podanie žaloby' (*Slovensko.sk: Ústredný portál verejnej správy (ÚPVS)*, 14 September 2018) <https://www.slovensko.sk/sk/zivotne-situacie/zivotna-situacia/_elektronicke-podanie-zaloby1> accessed 2 October 2025.

86 'Fyzické osoby, fyzické osoby – podnikatelia, právnické osoby: Videonávody' (*Slovensko.sk: Ústredný portál verejnej správy (ÚPVS)*, 25 February 2025) <https://www.slovensko.sk/sk/navody/videonavody> accessed 2 October 2025.

87 'Návody pre fyzické osoby' (*Slovensko.sk: Ústredný portál verejnej správy (ÚPVS)*, 2025) <https://www.slovensko.sk/sk/navody/navody-pre-fyzicke-osoby> accessed 2 October 2025.

88 Civil Procedure Code of the Slovak Republic No 160/2015 Coll 'Civilný sporový poriadok' (amended 1 July 2024) <https://www.epi.sk/zz/2015-160> accessed 2 October 2025.

89 Law of the Slovak Republic No 307 on Reminder Proceedings and on Amendments to Certain Acts 'O upomínacom konaní a o doplnení niektorých zákonov' (25 October 2016) <https://www.slov-lex.sk/ezbierky/pravne-predpisy/SK/ZZ/2016/307/> accessed 2 October 2025; Ministerstvo spravodlivosti Slovenskej republiky, *Používateľská príručka oblasti eŽaloby* (Rozvoj elektronických služieb súdnictva, Informačný systém elektronických služieb súdnictva 2018) 9, 19, 22 <https://obcan.justice.sk/documents/20229/0/EZA+Pr%C3%ADru%C4%8Dka+v14.5/77d634a7-4f29-4aa2-b8f7-7b315bd33f13> accessed 2 October 2025.

validation, and error warnings, with a total of 50 MB and an 8 MB per-file limit. Efficiency is supported by electronic forms, reusable templates, and integration with the Bar Association's information system.[90]

Slovenia. In 2024, Slovenia demonstrated strong ICT-based communication among courts, prosecutors, and legal professionals, but lacked chatbots and information on alternative legal aid providers. While procedural ICT use is established, blockchain and AI remain unimplemented. Case initiation and management – especially in administrative and criminal matters – rank among the lowest in the EU. Machine-readability standards are supported, though algorithmic anonymisation and pseudonymisation are absent.[91] In 2025, indicators remain almost unchanged.[92]

In 2023, Slovenia launched a national electronic identification scheme (e-ID) and an electronic signature system, forming the basis for e-justice through secure authentication and legally valid digital documents. Digital public service indicators rose from 71,4% to 77.0% for citizens and from 82,7% to 84.0% for businesses, yet a digital divide and low usage persist due to limited digital literacy.[93] The Unified Justice Portal (Portal e-Sodstvo) has simplified the procedure for submitting documents by ensuring their legal validity through a digital signature and enhancing transparency.[94] The overall accessibility of the system confirms an adequate level of development of digital governance in the field of justice in Slovenia.

The Slovenian Civil Procedure Act establishes equal legal force for electronic and written procedural documents, including their evidentiary value (Article 16a), and, in particular, for electronic public documents (Article 224). Applications are submitted electronically through the Portal e-Sodstvo information system, which provides automatic confirmation of receipt; the court generates either electronic or paper copies for delivery to the other party (Articles 105.b, 106). Applying electronically automatically implies consent to electronic delivery via the information system unless the party explicitly states otherwise. For state authorities, attorneys, and other individuals defined by law, documents are always delivered electronically (Article 132). Access to case materials for parties and case tracking is also provided via e-Sodstvo (Article 150). Remote hearings are permitted with the parties' consent (Article 114.a), during which the court may record the minutes using technical means, including audio or video recording (Articles 125, 125.a). Court decisions are produced electronically and signed with the electronic signature of the presiding judge (Article 323) and delivered electronically to the parties.

90 Ministerstvo spravodlivosti Slovenskej republik (n 89) 5, 7, 9, 13, 19, 30 56, 63, 92-3, 95.

91 European Commission, *The 2024 EU Justice Scoreboard* (n 35) 35-8.

92 European Commission, *The 2025 EU Justice Scoreboard* (n 32) 33-40.

93 European Commission, *Digital Decade Country Report 2024: Slovenia* (Publications Office of the EU 2024) 4, 7, 25 <https://ec.europa.eu/newsroom/dae/redirection/document/106696> accessed 2 October 2025.

94 *Portal e-Sodstvo* (*Republike Slovenija*) <https://evlozisce.sodisce.si/esodstvo/index.html> accessed 2 October 2025.

Court orders that are automatically processed within the information system are certified with an electronic seal (Article 329). Case transfers between instances (appeal, cassation, remand) are carried out electronically (Articles 345, 361, 375).[95] Thus, the Portal e-Sodstvo[96] functions as a centralised gateway for interaction with the judicial system. Access requires registration using an email address and password, while advanced services require "qualified user" status.[97] The system mandates digital signing of documents in PDF/A format using Adobe Acrobat Reader or LibreOffice to ensure integrity and legal validity.[98] The portal is available daily from 08:00 to 20:00 and supports compatibility with multiple operating systems (Windows, macOS, Linux).[99] The user guides provide detailed instructions on registration, password recovery, data modification, and account deletion, making the system accessible to a wide range of users.[100]

Central and Eastern European countries demonstrate a differentiated implementation of European standards, with a gradual transformation of justice systems. Thus, Poland achieves a balanced application of human-centricity, inclusivity, transparency, and security through multi-level authentication but faces gaps in informational accessibility and technological innovation. The Czech Republic and Slovakia implement access principles through ISDS/eŽaloby platforms with QES and promote human-centricity via inclusive instructions; however, their realisation of data-driven justice remains limited. Slovenia develops a human-centred e-Sodstvo Portal with multi-level support and effectively implements transparency and security principles. Bulgaria demonstrates the most ambitious implementation of the "digital by default" approach and achieves the highest indicators of data-driven justice, but faces challenges related to the digital divide. Hungary achieves high technical performance while facing systemic challenges to the principle of respect for fundamental rights. Croatia and Romania ensure human-centricity, transparency, and security through QES and GDPR compliance, but exhibit fragmented implementation. Common challenges across the region include the implementation of bridging the digital divide due to uneven complexity of access; limited enhancement of

95 Civil Procedure Act of the Republic of Slovenia of 25 March 1999 'Zakon o pravdnem postopku (ZPP)' (amended 25 October 2024) <https://pisrs.si/pregledPredpisa?id=ZAKO1212> accessed 2 October 2025.

96 *Portal e-Sodstvo* (n 94)

97 Vrhovno sodišče Republike Slovenija, 'Uporabniška navodila: E-Vloga V Civilnih Postopkih - Navodila za uporabo' (*Portal E-Sodstvo*, 12 February 2021) <https://evlozisce.sodisce.si/esodstvo/index.html> accessed 2 October 2025.

98 Vrhovno sodišče Republike Slovenija, 'Uporabniška navodila za registrirane uporabnike portala e-Sodstvo' (*Portal E-Sodstvo*, 29 March 2011) <https://evlozisce.sodisce.si/esodstvo/index.html> accessed 2 October 2025.

99 'Vrhovno sodišče Republike Slovenija, 'Uporabniška navodila: Navodila za namestitev podpisne komponente ProXSign' (*Portal E-Sodstvo*, 1 December 2022) <https://evlozisce.sodisce.si/esodstvo/index.html> accessed 2 October 2025.

100 Vrhovno sodišče Republike Slovenija, 'Uporabniška navodila: Navodila Podpisovanje PDF Dokumentov: Navodila za digitalno podpisovanje' (*Portal E-Sodstvo*, 2024) <https://evlozisce.sodisce.si/esodstvo/index.html> accessed 2 October 2025.

digital capabilities caused by the lack of innovative technologies and interactive tools for assessing legal aid; insufficient implementation of data-driven justice due to gaps in machine-readability; disruption of interoperability caused by fragmented systems; constraints on dynamic justice in criminal proceedings; and an ongoing imbalance between security and accessibility.

3.3. National Practices of Digital Justice in the Countries of the Baltic Region

Latvia. Latvia demonstrates a comprehensive and well-regulated digital judiciary, ranking among Europe's leaders. Electronic communication across judicial actors and robust ICT regulation in all types of proceedings ensure high system performance. However, digital tools for criminal cases remain limited, blockchain and AI are unused, and the machine-readability of court decisions is hindered by missing metadata and ECLI integration.[101] According to 2025 data, the availability of online legal aid has increased.[102]

Latvia exceeds the EU average in both public services for citizens (88.2% compared to 79.4% in the EU) and for businesses (87.2% compared to 85.4% in the EU). 78.4% of Latvians use e-government services, which is above the EU average. However, limited use of centralised applications and the once-only principle contribute to fragmented digital governance, affecting justice administration. Electronic document submission, party identification, and authentication are supported through e-ID, e-signature, and eSignature mobile systems.[103] The online platform E-lietas[104] provides electronic services in pre-trial and judicial proceedings as well as in the enforcement of decisions.[105] The platform operates exclusively with documents certified by a qualified electronic stamp, signature, and time label in accordance with EU Regulation No. 910/2014 (eIDAS).[106] The Court Administration controls personal data to organise access (Art. 9(1)).[107] Enforcement applications are submitted exclusively electronically (§ 149, § 150). Since 1 December 2021, an electronic signature has been mandatory for documents within the online system and for court decisions (§ 170, § 171). Electronic copies of paper documents have the same legal force as originals if standards are met, and courts have begun

101 European Commission, *The 2024 EU Justice Scoreboard* (n 35) 36-40.

102 European Commission, *The 2025 EU Justice Scoreboard* (n 32) 37.

103 European Commission, *Digital Decade Country Report 2024: Latvia* (Publications Office of the EU 2024) 22-4 <https://ec.europa.eu/newsroom/dae/redirection/document/106714> accessed 4 October 2025.

104 *E-lietas portāls (Latvijas Republika)* <https://elieta.lv> accessed 4 October 2025.

105 Law of the Republic of Latvia on the State Platform for Electronic Proceedings 'Procesu norises elektroniskā vidē valsts platformas likums' (10 March 2022) <https://likumi.lv/ta/id/330962-procesu-norises-elektroniska-vide-valsts-platformas-likums> accessed 4 October 2025.

106 Civil Procedure Law of the Republic of Latvia of 14 October 1998 'Civilprocesa likums' (amended 1 April 2025) <https://likumi.lv/ta/id/50500-civilprocesa-likums> accessed 4 October 2025.

107 Regulations of the Cabinet of Ministers of the Republic of Latvia No 217 'E-case Platform Data Processing Regulations' (5 April 2022) <https://likumi.lv/ta/id/331413-e-lietas-platformas-datu-apstrades-noteikumi0> accessed 4 October 2025.

converting paper-based records into electronic form (§ 172, § 167). For electronic communication, the claimant provides consent and an electronic address or indicates registration in the online system (§ 128(2)). Representatives outside Latvia may only provide an email address or indicate registration (§ 128(2)).[108]

Estonia. Estonia, according to the 2024 EU Justice Scoreboard, demonstrates the highest indicators of digitalisation in justice across the EU. Areas identified for improvement include enhancing the machine-readability of court decisions, in particular by implementing decision modelling in accordance with relevant standards.[109] The 2025 EU Justice Scoreboard highlights the absence of interactive rights training, online legal aid eligibility assessment, and informational chatbots in Estonia. Nonetheless, Estonia leads the EU in producing machine-readable court decisions.[110]

Estonia leads the EU in digital public services, with 95.8% coverage for citizens and 98.8% for businesses. Over 80% of the population holds an e-ID, and 89% use it for online services, far exceeding EU averages. National law requires all citizens over 15 years old to possess an EU-recognised e-ID, ensuring secure identification and data protection. Estonia also outperforms the EU in digital skills and cybersecurity, reinforcing digital inclusion and a human-centred digital transformation.[111]

The e-Toimik portal[112] serves as Estonia's central electronic justice system for managing procedural data, handling electronic case files, and facilitating interaction with the courts (§ 60^1 (1) of the Estonian Code of Civil Procedure). The e-Toimik portal ensures a full cycle of digital justice, including case review, court workflow organisation, statistical reporting, and electronic document transmission (§ 60^1). The Ministry determines the procedures for transitioning to mandatory digital case management and establishes the relevant technical requirements (§ 61). Digital case files are handled with chronological storage of documents from all court instances. Paper documents are automatically scanned and substituted by their digital versions, with the originals kept temporarily until the decision becomes final (§ 56). In cases of technical malfunctions, paper case files may be temporarily handled and subsequently converted into digital form (§ 57^1). Participants have full access to their case files, with the option to obtain electronic copies or printed versions (§ 59). Electronic delivery of documents is carried out via the information system, with notifications sent to email addresses, mobile numbers, and social media accounts. A document is considered delivered upon opening or confirmation, with automatic registration (§ 311^1). For legal professionals (including attorneys, notaries, bailiffs, and government bodies), electronic delivery is mandatory; failure to comply may

108 Civil Procedure Law of the Republic of Latvia (n 106).

109 European Commission, *The 2024 EU Justice Scoreboard* (n 35) 36-40.

110 European Commission, *The 2025 EU Justice Scoreboard* (n 32) 33.

111 European Commission, *Digital Decade Country Report 2024: Estonia* (Publications Office of the EU 2024) 20-1 <https://ec.europa.eu/newsroom/dae/redirection/document/106705> accessed 4 October 2025.

112 *E-toimik (Eesti Vabariik)* <https://www.e-toimik.ee> accessed 4 October 2025.

result in temporary restrictions on access to integrated systems (§ 311[1]). Electronic documents are submitted with a digital signature or another sender authentication method. However, the court may accept unsigned documents if there is no doubt about their authenticity and digital copies are generated automatically (§ 336, § 340).[113]

Lithuania. The 2025 EU Justice Scoreboard shows that Lithuania performs strongly in judicial digitalisation and offers extensive online information on the justice system. However, chatbots, AI, and blockchain remain unimplemented, and defendants in criminal cases cannot apply for legal aid online.[114] Compared to the 2024 Scoreboard, there are no significant dynamics, which is generally explained by Lithuania's already high performance indicators.[115]

The digitalisation of public services in Lithuania exceeds the EU average, both for citizens (86.7% compared to 79.4% in the EU) and for businesses (95.9% compared to 85.4% in the EU), with an ambition to reach 100% by 2030. Lithuania is a leader in the use of e-ID. An updated version of the Law on the Management of State Information Resources has been adopted to enhance governance and interoperability of digital systems. At the same time, the use of AI remains limited, and 88% of Lithuanians believe that public authorities should provide better support to help citizens adapt to digital technologies.[116]

In Lithuania, the Lithuanian Court E-Services Portal (LITEKO) serves as the central platform for judicial digitalisation.[117] Electronic technologies are extensively integrated into civil proceedings, including the automated allocation of cases without altering jurisdiction (Articles 622, 622-3 of the Code of Civil Procedure), the conduct of court hearings, the exchange of documents, and the remote participation of parties via video or teleconferencing, either using court-provided tools or personal devices, except in written procedures (Article 175-2). Upon a reasoned request or when necessary, the court may schedule an in-person hearing. An electronic signature is mandatory for all electronic documents. The exchange, delivery of documents, and summons are carried out through LITEKO, with mandatory notifications sent to the participants' email addresses, phone numbers, and fax numbers. For legal professionals, including attorneys, assistants, bailiffs, and notaries, delivery is conducted exclusively electronically (Articles 175-1, 113, 122), with the date of delivery deemed to be the following working day. In cases involving a court order,

113 Code of Civil Procedure of the Republic of Estonia of 20 April 2005 'Tsiviilkohtumenetluse seadustik' (amended 13 April 2025) <https://www.riigiteataja.ee/akt/104072012019?leiaKehtiv> accessed 4 October 2025.

114 European Commission, *The 2025 EU Justice Scoreboard* (n 32) 34, 37-8.

115 European Commission, *The 2024 EU Justice Scoreboard* (n 35) 33-40.

116 European Commission, *Digital Decade Country Report 2024: Lithuania* (Publications Office of the EU 2024) 4, 18, 22 <https://ec.europa.eu/newsroom/dae/redirection/document/106704> accessed 4 October 2025.

117 *Lietuvos teismų elektroninių paslaugų portalas* <https://e.teismas.lt/en/public/home/> accessed 4 October 2025.

the applicant must receive the documents electronically (Article 432-1).[118] An order of the Ministry of Justice regulates the digital procedures.[119]

The Baltic countries most comprehensively implement the European paradigm through the integrated application of the principles of the Declaration of Digital Rights and Principles for the Digital Decade (2023) and the European e-Justice Strategy 2024–2028. Estonia demonstrates exemplary implementation of human-centricity through universal e-ID, inclusivity through widespread digital literacy, "digital by default" with mandatory electronic procedures for professionals, and a "fair digital environment" through open technologies. Security and privacy principles are ensured by reliable authentication within the e-Toimik system, interoperability is achieved through the full integration of judicial systems, and access to justice is facilitated by the 95.8% digitalisation of public services. Latvia effectively realises the principles of "once-only" and dynamic justice through the E-lietas platform with automated case distribution, exceeding EU average performance indicators. Lithuania implements efficiency principles through the LITEKO system with automated routing. Shared challenges include the incomplete implementation of data-driven justice due to gaps in machine-readability; constraints on dynamic justice in criminal proceedings; the need to strengthen inclusivity due to the lack of chatbots and online training; and the limited implementation of innovative digital justice resulting from the lack of AI and blockchain technologies in core activities, which remains acceptable at the current stage considering the gradual introduction of the AI Act.

The analysis of the digitalisation of justice in the selected countries enables us to identify the approaches that meet the stated criteria. Systemic integration (Estonia, Latvia, Lithuania) combines universal mandatory electronic filing with unified interoperable platforms, achieving stronger performance across the evaluation indicators. Gradual transformation (Poland, Czechia, Slovakia) mandates electronic filing for legal professionals while retaining traditional channels, supported by partially integrated systems with moderate levels of machine-readability and human-centred design. Implementation, notably characteristic of Croatia, Romania, and Bulgaria, is technologically advanced; however, the portals are not integrated, voluntary electronic filing is inconsistent, and gaps persist in platform integration and the quality of user guidance, despite certain advantages (for example, machine-readable decisions in Bulgaria). The comparative analysis of the criteria for digital justice quality reveals divergences among countries. Estonia and Bulgaria employ XML/JSON formats with ECLI identifiers, whereas Poland, Czechia, and Croatia rely on

118 Civil Procedure Code of the Republic of Lithuania No IX-743 of 28 February 2002 'Civilinio proceso kodeksas' (amended 31 August 2025) <https://e-seimas.lrs.lt/portal/legalAct/lt/TAD/TAIS.162435/asr> accessed 4 October 2025.

119 Order of the Minister of Justice of the Republic of Lithuania No 1R-332 'On Approval of the Procedure for Submission of Procedural Documents to the Court and Their Service to Persons by Electronic Means' (13 December 2012) <https://www.e-tar.lt/portal/lt/legalAct/TAR.A527F4DF2B60/asr> accessed 4 October 2025.

PDF formats. Slovakia's eŽaloby, Czechia's ePodatelna, and Slovenia's e-Sodstvo exhibit a human-centred design through the provision of video instructions and reusable templates, in contrast to Croatia's user manual and Romania's access limitations. Platform integration is most advanced in Estonia and Lithuania due to universal digital case processing and automated routing. By contrast, Croatia and Romania exhibit fragmented architectures and multiple parallel portals. A clearer distinction emerges between jurisdictions with universal mandatory electronic filing, exemplified by Estonia and Latvia, and those that rely on voluntary systems. Common elements include the universal adoption of qualified electronic signatures (in line with the eIDAS Regulation 910/2014), the use of videoconferencing, and mandatory electronic service for legal professionals. The corresponding data provide evaluation criteria for Ukraine, which is pursuing systemic reforms of its justice system.

3.4. Ukraine's Digital Justice Landscape: Challenges, Gaps and Alignment with EU Standards

Legal literature emphasises that the judiciary's efficiency constitutes the foundation of the contemporary rule of law, compelling Ukraine to comply with its international obligations under the Association Agreement between Ukraine and the EU (2014).[120] Thus, within the framework of fulfilling obligations under the Association Agreement, Ukraine is carrying out a phased adaptation of national legislation to the EU acquis in the field of digital transformation, including the digitalisation of justice.[121] According to the results of the first stage of the legislative screening (2023), Ukraine identified the need to bring some regulatory acts into compliance with European standards in the areas of electronic identification, cross-border data exchange, digital public services, and e-governance, which are directly related to the development of e-justice,[122] while the inclusion of these priorities in the Rule of Law Roadmap (2025) became an important step towards the implementation of the acquis EU, particularly in terms of personal data protection and the resilience of judicial infrastructure.[123] Tsuvina (2025) notes that Ukraine's integration of AI into the judiciary lacks the conceptual and regulatory maturity required under the emerging EU

120 Iryna Izarova and others, 'Advancing Sustainable Justice through AI-Based Case-Law Analysis: Ukrainian experience' (2024) 7(1) Access to Justice in Eastern Europe 127. doi:10.33327/AJEE-18-7.1-a000123.

121 Association Agreement between the European Union and the European Atomic Energy Community and their Member States, of the one part, and Ukraine, of the other part (signed 27 June 2014, effective 1 September 2017) <https://www.consilium.europa.eu/en/documents/treaties-agreements/agreement/?id=2014045> accessed 8 October 2025.

122 Cabinet of Ministers of Ukraine, *Report on the Initial Assessment of the Progress in the Implementation of the European Union Legal Acts (EU Acquis)* (European Union 2023) <https://eu-ua.kmu.gov.ua/wp-content/uploads/Zvit_EN.pdf?> accessed 8 October 2025.

123 Cabinet of Ministers of Ukraine, 'Roadmap on the Rule of Law' (*Ministry of Justice of Ukraine*, 14 May 2025) <https://minjust.gov.ua/news/ministry/ukraina-zatverdila-dorojni-karti-v-mejah-vstupu-do-es-opublikovano-teksti-dokumentiv> accessed 8 October 2025.25

acquis. Although pilot initiatives exist, the absence of clear safeguards, risk classification, and accountability mechanisms places Ukraine in an early pre-compliance stage with respect to EU AI governance.[124]

Digital transformation of judicial proceedings in Ukraine can be characterised by the implementation of the Law of Ukraine "On Access to Court Decisions"[125] and the "Procedure for Maintaining the Unified State Register of Court Decisions".[126] The Unified Judicial Information and Telecommunication System (UJITS)[127] in Ukraine has been officially functioning since 5 October 2021 and was introduced gradually[128]. Access to information about the judiciary in Ukraine, including its components,[129] the procedure for using the UJITS, the submission of procedural documents through the "Electronic Court" subsystem for citizens and organisations, and user instructions, is provided in an accessible form.[130]

Ukraine, like other European states, accelerated the implementation of the electronic court system during the COVID-19 pandemic. The concept was formulated as early as 2012 and was legally established in 2018 through the creation of UJITS. Relevant amendments were introduced into all procedural codes. In 2021, the full functioning of the subsystems "Electronic Court," "Electronic Cabinet," and videoconferencing was officially announced, along with the presentation of the mobile application "Court in a Smartphone," designed to ensure full access to e-court services via mobile devices.[131] Although users consistently reported instability in its operation.[132]

124 Tetiana Tsuvina, 'Artificial Intelligence Technologies in the Judiciary: European Standards and Ukrainian Practice' (2025) 139(2) Foreign Trade: Economics, Finance, Law 4. doi:10.31617/3.2025(139)03.

125 Law of Ukraine No 3262-IV of 22 December 2005 'On Access to Court Decisions' (amended 25 March 2025) <https://zakon.rada.gov.ua/laws/show/3262-15#Text> accessed 8 October 2025.

126 Decision of the High Council of Justice No 1200/0/15-18 'On Approval of the Procedure for Maintaining the Unified State Register of Court Decisions' (19 April 2018) <https://zakon.rada.gov.ua/go/v1200910-18> accessed 8 October 2025.

127 Law of Ukraine No 1416-IX of 27 April 2021 'On Amendments to Certain Legislative Acts of Ukraine Regarding Ensuring the Phased Implementation of the Unified Judicial Information and Telecommunication System' [2021] Official Gazette of Ukraine 42/2502.

128 Law of Ukraine No 3200-IX of 29 June 2023 'On Amendments to Certain Legislative Acts of Ukraine Regarding Mandatory Registration and Use of Electronic Cabinets in the Unified Judicial Information and Telecommunication System or its Separate Subsystem (Module) that Ensures Document Exchange' [2023] Official Gazette of Ukraine 70/4041.

129 *Website of the Supreme Court of Ukraine* <https://court.gov.ua> accessed 8 October 2025.

130 'Functioning of the Unified Judicial Information and Communication System' (*State Judicial Administration of Ukraine*, 2025) <https://dsa.court.gov.ua/dsa/inshe/func_ecits/> accessed 8 October 2025.

131 Olexander P Svitlychnyy and others, 'Electronic Justice as a Mechanism for Ensuring the Right of Access to Justice in a Pandemic: The Experience of Ukraine and the EU' (2023) 37(3) International Review of Law, Computers & Technology 325. doi:10.1080/13600869.2023.2221820.

132 'eCourt Mobile Application' (*Google Play Store*, 24 May 2023) <https://play.google.com/store/apps/details?id=com.floor12apps.ecourt&hl=uk> accessed 8 October 2025.

Kaniuka (2023) similarly observes that Ukraine's e-justice ecosystem, while normatively ambitious, remains technologically inconsistent and structurally under-integrated. She identifies recurring problems, including uneven digital readiness across courts, limited interoperability between registries, and insufficient user support, all of which hinder the realisation of EU-level standards of accessibility and transparency.[133]

It is also worth supporting the position of Maika (2022), who argues that electronic access to justice is an urgent need that requires improving regulatory frameworks and ensuring adequate technical and organisational preconditions: establishing procedures for submitting and evaluating electronic evidence; standardising and certifying IT infrastructure within the e-justice system; developing legal education in the field of e-governance; and enhancing the digital literacy of citizens and public officials.[134]

At present, Ukraine has a solid foundation and an ambitious concept for improving its e-justice, but the existing infrastructure still requires significant modernisation to meet European standards. According to the Rule of Law Roadmap (2025), during 2026–2027, several key initiatives are planned to enhance access to justice, transparency, and the efficiency of courts through digital transformation. These include the development and approval of a comprehensive Concept and Roadmap for the creation of IT solutions within the judicial system, the adoption of a financing plan with clearly defined funding sources and institutional responsibilities, and the establishment of an effective management framework to oversee IT reforms and support within the justice (p. 18). Moreover, Ukraine intends to adopt new legislation to harmonise national regulations with the EU acquis regarding the digitalisation of justice within the European Union and to implement the necessary IT solutions in line with the European e-Justice Strategy for 2024–2028.[135]

The Civil Procedure Code of Ukraine regulates the use of digital tools for accessing and conducting civil proceedings via the UJITS. It establishes mandatory automated registration of documents (Art. 14(2)) and automated case distribution (Art. 14(3)), ensuring full electronic document flow between courts and participants (Art. 14(4)), including video conferencing and recording of court hearings. A differentiated approach to registering electronic accounts is provided: mandatory for professional participants in judicial proceedings and state authorities (Art. 14(6)), with procedural sanctions for non-compliance (Arts. 117(4), 153(10)), and voluntary for other individuals. Electronic documentation is carried out through UJITS, which requires a mandatory qualified electronic signature (QES) (Art. 43(8)) and requires proof that copies of documents were sent to other participants (Art. 43(7)), thereby safeguarding the principle of adversarial

133 Natalia Kaniuka, 'Trends and Problems of e-Justice Enforcement in Ukraine' (2023) 1 Visegrad Journal on Human Rights 95. doi:10.61345/1339-7915.2023.1.13.

134 Maika Maksym, 'The Implementation of E-Justice within the Framework of the Right to a Fair Trial in Ukraine: Problems and Prospects' (2022) 5(3) Access to Justice in Eastern Europe 249. doi:10.33327/AJEE-18-5.2-n000320.

135 Cabinet of Ministers of Ukraine, 'Roadmap on the Rule of Law' (n 123).

proceedings in a digital environment. The legislator maintains a transitional approach, allowing mixed document management (Art. 14(8)). Remote participation in court hearings is regulated concerning the specifics of the parties: general opportunity for parties and their representatives to participate via video conference outside the courtroom (Art. 212(1)) and limited participation for procedural assistants, who may participate exclusively within court premises (Art. 212(6)). Court decisions and procedural documents are created in electronic form and signed with a qualified electronic signature (Art. 259(8)), ensuring the legal validity of digital acts. The UJITS security system provides protection for information and establishes legal liability for unauthorised interference (Arts. 14(11–12)), thereby meeting the minimum requirements for digital security in judicial proceedings.[136]

In Ukraine, a clear strategy or conceptual framework for justice sector reform in the context of digitalisation has not yet been articulated, and the criteria for implementation and their evaluation remain undefined and, overall, not institutionalised. Consistent and transparent monitoring has not been a firmly established element of Ukraine's legal tradition. Considering this, the vectors of Ukrainian justice reforms and the assessment of the state of digitalisation of justice may be presented—and, in the future, conducted—using the indicators proposed by the EU Justice Scoreboard. Ukraine demonstrates strong progress in several areas: the extensive use of automated document management systems across all courts, electronic court decisions, a centralised automatic case allocation system, integration with state registers, widespread use of electronic signatures, and the provision of electronic offices for all professional participants. The system also offers automatic notifications about procedural actions, integration with the Automated Enforcement Proceedings System, and basic-level video conferencing, including dedicated functionality for courtrooms, detention centres, correctional facilities, and medical institutions. Additionally, remote participation is enabled for various types of proceedings. However, significant challenges remain. The lack of structured metadata and machine-readable standards (e.g. XML/JSON formats, semantic tagging of elements, ECLI identifiers, automated extraction of legal norms from texts, linking cases to decisions, multilingualism, automated translation, and open APIs for developers and researchers) limits efficiency and transparency. Other shortcomings include outdated interface design, confusing navigation, and the necessity to work simultaneously with several disparate systems (Unified State Register of Court Decisions (USRCD), electronic offices, video conferencing tools, and registers). There is also a lack of user guidance, lengthy registration and authorisation procedures, and service fragmentation across multiple subsystems. Moreover, Blockchain/DLT technologies are neither implemented nor foreseen in current strategic planning. Effective implementation of EU e-justice standards in Ukraine requires the consistent integration of innovative technologies and the establishment of a secure electronic communication system. A key determinant of this

136 Civil Procedure Code of Ukraine No 1618-IV of 18 March 2004 (amended 17 July 2025) <https://zakon.rada.gov.ua/laws/show/1618-15> accessed 8 October 2025.

process's effectiveness is maintaining a balance between digital innovation, the protection of fundamental rights and personal data, and the resilience of the electronic judicial infrastructure.[137] However, achieving interoperability with EU systems remains a significant challenge. Ukraine lacks standardised APIs and shared data protocols necessary for intersystem integration. Court decisions are not presented in machine-readable XML/JSON formats, which limits automated processing. The absence of ECLI identifiers further impedes integration with EU judicial networks.

The challenges faced by Ukraine mirror those identified by Ramos Maqueda and Chen (2025) in jurisdictions that have not transitioned to data-driven justice. In the absence of structured datasets, stable identifiers, and interoperable information architectures, digitalisation remains superficial and cannot support advanced analytics, automation, or evidence-based judicial governance.[138] Razmetaeva and Razmetaev (2021) warn that such infrastructural fragmentation poses bigger constitutional risks, as digital uncertainty and opacity may undermine legal certainty and equality of arms. They emphasise that Ukraine's digitalisation must advance through both technological consolidation and robust protection of procedural rights.[139]

The absence of a comprehensive digital justice strategy and action plan, alongside the ambitious UJITS concept, creates a paradoxical situation: while funding for UJITS reform (which encompasses more than ten subsystems and modules) has been suspended, professional discussions increasingly focus on AI implementation[140] without an appropriate legal framework or strategic vision. This reflects Ukraine's fragmented approach to e-justice development, where advanced technologies are discussed without addressing fundamental infrastructure gaps. There is a discrepancy between European standards on the use of AI in justice and the current understanding of the application of such technologies in professional activities,[141] without taking into account the EU AI Act.[142] The Code of Judicial Ethics only provides that the use of AI by a judge is permissible, provided that judicial independence

137 Oleh Syniehubov, Oksana Bortnik and Olena Chernenko, 'Strategic Goals and Principles of Digitalisation of Justice in the EU in the Context of European Integration Challenges for Ukraine' (2024) 5 Law Herald 184. doi:10.32782/yuv.v5.2024.21.

138 Manuel Ramos-Maqueda and Daniel L Chen, 'The Data Revolution in Justice' (2025) 186 World Development 106834. doi:10.1016/j.worlddev.2024.106834.

139 Razmetaeva and Razmetaev (n 33).

140 Order of the Cabinet of Ministers of Ukraine No 1556-p 'On the Approval of the Concept for the Development of Artificial Intelligence in Ukraine' (2 December 2020) <https://zakon.rada.gov.ua/laws/show/1556-2020-%D1%80#n8> accessed 8 October 2025; Order of the Cabinet of Ministers of Ukraine No 457-p 'On the Approval of the Action Plan for the Implementation of the Concept for the Development of Artificial Intelligence in Ukraine for 2025–2026' (9 May 2025) [2025] Official Gazette of Ukraine 45/3085.

141 Yan Bernaziuk, 'Artificial Intelligence and Ukraine's Justice System: Results of Cooperation Over the Past Year' (*Supreme Observer*, 7 January 2025) <https://so.supreme.court.gov.ua/authors/934/shtuchnyi-intelekt-ta-systema-pravosuddia-ukrainy-rezultaty-spivpratsi-u-rotsi-sh%D1%81ho-mynuv> accessed 8 October 2025.

142 Regulation (EU) 2024/1689 (n 6).

and impartiality are maintained, that it does not relate to the assessment of evidence or the decision-making process, and that it does not violate legislative requirements (Art. 16).[143] The concept of the UJICS indicates the lack of details on AI control mechanisms and lacks a clear list of prohibitions; nevertheless, the approach remains responsible and considers AI solely in an assistive role within judicial systems. Limited functions and specificity characterise the concept through a precise list of permitted uses, including OCR, speech-to-text and text-to-speech technologies, translation, grammar checking, case law search, generation of draft documents, and virtual assistant avatars. Quality control functions include identifying deviations during court proceedings or inconsistencies with general judicial practice, as well as providing basic legal assistance to users. Transparency in accordance with the concept that reflects European standards must be ensured through mandatory system self-learning and the preservation of human control over all decisions.[144]

Further implementation of EU e-justice standards requires a comprehensive approach. Technical modernisation must include not only the transition to machine-readable formats but also the creation of unified data-exchange protocols. Ensuring full interoperability requires integrating fragmented systems into a single ecosystem using semantic technologies. It is critically important to enhance the user experience by simplifying procedures and developing an inclusive design for vulnerable groups. Regulatory harmonisation should encompass the full spectrum of EU requirements—from eIDAS to cross-border data exchange. Successful transformation is possible only if technological innovation is balanced with the protection of human rights and the resilience of judicial infrastructure, while learning from both successful implementations (Estonia, Latvia) and the fragmentation challenges (Romania, Croatia) observed across EU Member States.

4 CONCLUSIONS

The European approach to digital justice demonstrates a systemic shift in the understanding of technology's role, from tools aimed at increasing efficiency to instruments for strengthening fundamental rights. Unlike approaches to technological determinism, the EU establishes a value-oriented digitalisation process in which each technical solution is subordinated to legal principles and guarantees. EU's normative basis ensures the protection of procedural rights in digital environments, while value-oriented instruments reinforce legitimacy by framing digital justice as inclusive, human-centred, and rights-compliant. Strategic frameworks outline long-term priorities, supported by operational tools for implementation and secure cross-border cooperation. Institutional coherence is achieved

143 Decision of the Congress of Judges of Ukraine 'On the Approval of the Code of Judicial Ethics' (18 September 2024) <https://zakon.rada.gov.ua/go/n0001415-24> accessed 8 October 2025.

144 Order of the State Judicial Administration of Ukraine No 178 'Concept of the Unified Judicial Information and Telecommunication System' (30 April 2025) <https://court.gov.ua/storage/portal/dsa/normatyvno-pravova%20baza/N_178_2025_dodatok.pdf> accessed 8 October 2025.

through methodological recommendations focused on efficiency, transparency, and inclusiveness, while progress is monitored via performance assessments and comparative indicators. Together, these instruments establish a structured and layered governance model that integrates soft and hard law, ethical values, and technical standards, and enables both EU-wide coordination and national adaptation in the field of digital justice. An analysis of the European legal acts underpinning digital transformation and national practices reveals the distinctive nature of this approach: digital rights are directly integrated into the architecture of judicial systems through specific technical standards, procedural requirements, and prohibitions on fully automated decision-making.

A study of civil procedural legislation and e-justice platforms in Central and Eastern Europe and the Baltic states reveals consistent patterns in the implementation of European standards, confirming the European paradigm's flexibility while maintaining shared value-based foundations. The Baltic states are developing a regional strategy for systemic integration with high levels of automation and mandatory electronic document management, with Estonia achieving the best results. The approach of Central and Eastern European countries is more differentiated. The experiences of Poland, the Czech Republic, and Slovakia illustrate a gradual e-transformation that balances digital and traditional access channels. The Polish practice of mandatory electronic delivery for professional participants, while retaining alternative channels for citizens, demonstrates the possibilities of a phased transition to digital procedures. The Bulgarian approach represents a radical transformation: the shift of many civil cases to an electronic format with centralised allocation from July 2025 contrasts sharply with evolutionary approaches. The Hungarian case highlights the complexity of the interaction between technological and institutional contexts: high levels of digitalisation coexist with systemic challenges to democratic principles. The Croatian and Romanian experiences reveal fragmentation issues, where technologically advanced solutions fail to integrate into a unified judicial ecosystem.

The analysis identifies an important trend: machine-readable standards for court decisions are a critical prerequisite for achieving data-driven justice. Structured metadata, semantic markup, and API compatibility create a technological foundation for judicial analytics, predictive modelling, and legal assistance. Machine-readability is not merely a technical characteristic but forms the basis of the justice of the future. Data from the EU Justice Scoreboard 2024–2025 and national e-platform analyses confirm steady progress in digital transformation. At the same time, an examination of user instructions highlights significant differences in accessibility, ranging from user-friendly interfaces to multi-step authentication procedures, which pose risks to equitable access to justice.

Ukraine finds itself in a unique position on the path to European integration and the digitalisation of justice. Despite an ambitious UJITS concept and wartime experience that accelerated digitalisation, the country still lags behind European standards of inclusivity and system integration. While Ukraine has adopted certain digital mechanisms—such as electronic court services and remote hearings—these remain fragmented and unevenly

applied. Gaps persist in institutional interoperability, legal certainty, digital accessibility, and the protection of vulnerable groups. Challenges remain in implementing the "digital by default" principle due to process duplication. The primary risk lies in focusing on technical excellence while neglecting real user needs. The European approach of e-justice illustrates the evolution of access to justice from merely removing barriers to actively enabling opportunities through digital education, technical support, and intuitive interfaces. The principle of technological subsidiarity becomes a critical element of judicial systems: technologies perform an assistive role rather than replacing human judgment, ensuring transparency in automated processes. Estonia's experience offers an alternative strategy, emphasising simplified interaction with the system and creating positive incentives for a voluntary transition to digital procedures.

The findings suggest that Ukraine's digital justice agenda must adopt a systemic and standards-based approach. The comparative analysis of the selected jurisdictions reveals important lessons and data-driven priorities for the development of digital justice in Ukraine. Drawing on the state of normative regulation, technical infrastructure, and the assessment of the functioning of digital justice, the following directions may be identified which, within the framework of European integration processes, require particular attention beyond technical enhancement: the introduction of machine-readable formats for court decisions (XML/JSON and, in the longer term, with ECLI identifiers), since structured data enable more effective and automated legal analysis; the prioritisation of user-centred design through video tutorials, reusable templates, and intuitive interfaces, as successfully implemented in Slovakia and Czechia, which significantly reduces user barriers; the establishment of mandatory electronic filing for justice sector professionals, public authorities, and legal entities, while maintaining alternative channels for citizens in line with a balanced approach, such as that applied in Poland, which achieves digitalisation without exclusion; the consolidation of fragmented subsystems (USRCD, Electronic Court, videoconferencing, and others) into a unified API-compatible ecosystem, as the contrast between the integrated system in Estonia and the fragmentation in Croatia and Romania demonstrates that isolated technological sophistication without integration does not ensure comprehensive quality and efficiency; and the development of a legal framework for transparent AI governance in accordance with the requirements of the EU Artificial Intelligence Act, with a clear distinction between assistive systems and high-risk decision-support tools, particularly in view of the regulatory gap in Ukraine. These priorities, derived from comparative findings in Central European and Baltic countries, should be implemented through a unified governance structure that avoids ministerial fragmentation and is supported by institutionalised monitoring mechanisms and participatory policy-making. Success requires balancing institutional efficiency with citizen accessibility, ensuring a digital transformation that enhances rather than restricts access to justice for all population groups.

The research demonstrates that successful e-justice transformation requires balancing institutional efficiency with citizen accessibility. For Ukrainian citizens, this means ensuring that digital services do not exclude vulnerable populations while improving access to justice. For institutions, it necessitates comprehensive staff training, adequate funding, and sustained political commitment to avoid fragmentation, inconsistency, lack of transparency, and lack of accountability.

REFERENCES

1. Bernaziuk Y, 'Artificial Intelligence and Ukraine's Justice System: Results of Cooperation Over the Past Year' (*Supreme Observer*, 7 January 2025) <https://so.supreme.court.gov.ua/authors/934/shtuchnyi-intelekt-ta-systema-pravosuddia-ukrainy-rezultaty-spivpratsi-u-rotsi-sh%D1%81ho-mynuv> accessed 8 October 2025
2. Borgesano F and others, 'Artificial intelligence and justice: a systematic literature review and future research perspectives on Justice 5.0' (2025) 28(11) European Journal of Innovation Management 349. doi:10.1108/EJIM-01-2025-0117
3. Bułach R, 'E-doręczenia w administracji publicznej' (*Wolters Kluwer*, 13 August 2024) <https://www.wolterskluwer.com/pl-pl/expert-insights/e-doreczenia-w-administracji-publicznej> accessed 28 September 2025
4. Derhachova H and Koleshnia Y, 'Digital Business Transformation: Essence, Signs, Requirements and Technologies' (2020) 17 Economic Bulletin of National Technical University of Ukraine "Kyiv Polytechnical Institute" 280. doi:10.20535/2307-5651.17.2020.216367
5. Elia G and others, 'The Digital Transformation Canvas: A Conceptual Framework for Leading the Digital Transformation Process' (2024) 67(4) Business Horizons 381. doi:10.1016/j.bushor.2024.03.007
6. Fabri M, 'From Court Automation to e-Justice and Beyond in Europe' (2024) 15(3) International Journal for Court Administration 7. doi:10.36745/ijca.640
7. Fierbinţeanu G and Nemeş V, 'Digital Tools for Judicial Cooperation Across the EU – The Benefits of Digital Technologies in Judicial Proceedings' (2022) 15(1) Challenges of the Knowledge Society 136
8. Gascón Inchausti F, 'The New Regulation on the Digitalisation of Judicial Cooperation in the European Union: Something Old, Something New, Something Borrowed and Something Blue' (2023) 24 ERA Forum 535. doi:10.1007/s12027-024-00782-z
9. Gesley J, 'FALQs: The Rule of Law in the European Union' (*Library of Congress blogs*, 12 August 2022) <https://blogs.loc.gov/law/2022/08/falqs-the-rule-of-law-in-the-european-union/> accessed 25 September 2025.

10. Izarova I and others, 'Advancing Sustainable Justice through AI-Based Case-Law Analysis: Ukrainian Experience' (2024) 7(1) Access to Justice in Eastern Europe 127. doi:10.33327/AJEE-18-7.1-a000123

11. Kaniuka N, 'Trends and Problems of e-Justice Enforcement in Ukraine' (2023) 1 Visegrad Journal on Human Rights 95. doi:10.61345/1339-7915.2023.1.13

12. Makauskaite-Samuole G, 'Transparency in the Labyrinths of the EU AI Act: Smart or Disbalanced?' (2025) 8(2) Access to Justice in Eastern Europe 38. doi:10.33327/ajee-18-8.2-a000105

13. Maika M, 'The Implementation of E-Justice within the Framework of the Right to a Fair Trial in Ukraine: Problems and Prospects' (2022) 5(3) Access to Justice in Eastern Europe 249. doi:10.33327/AJEE-18-5.2-n000320

14. Nadkarni S and Prügl R, 'Digital Transformation: A Review, Synthesis and Opportunities for Future Research' (2021) 71 Management Review Quarterly 233. doi:10.1007/s11301-020-00185-7

15. Piatyhora KV, 'E-justice in Administrative Process: European Standards and Foreign Experience' (2024) 2(24) Theory and Practice of Jurisprudence 39. doi:10.21564/2225-6555.2023.2.293064

16. Piatyhora KV, 'International Experience of Implementing E-Justice: Best Practices and Challenges' (2024) 1(25) Theory and Practice of Jurisprudence 107. doi:10.21564/2225-6555.2024.1(25).305796

17. Pyrohovska V and others, 'E-Justice and the Development of Justice: Strengths, Challenges and Prospects' (2024) 16(1) Lex Humana 426

18. Ramos-Maqueda M and Chen DL, 'The Data Revolution in Justice' (2025) 186 World Development 106834. doi:10.1016/j.worlddev.2024.106834

19. Razmetaeva Y and Razmetaev S, 'Justice in the Digital Age: Technological Solutions, Hidden Threats and Enticing Opportunities' (2021) 4(2) Access to Justice in Eastern Europe 104. doi:10.33327/AJEE-18-4.2-a000061

20. Saprykin V, 'Digitization, Digitalization and Digital Transformation of Public Administration in Ukraine' (2024) 19(1) Bulletin of Taras Shevchenko National University of Kyiv: Public Administration 116. doi:10.17721/2616-9193.2024/19-19/22

21. Svitlychnyy OP and others, 'Electronic Justice as a Mechanism for Ensuring the Right of Access to Justice in a Pandemic: The Experience of Ukraine and the EU' (2023) 37(3) International Review of Law, Computers & Technology 325. doi:10.1080/13600869.2023.2221820

22. Syniehubov O, Bortnik O and Chernenko O, 'Strategic Goals and Principles of Digitalisation of Justice in the EU in the Context of European Integration Challenges for Ukraine' (2024) 5 Law Herald 184. doi:10.32782/yuv.v5.2024.21

23. Tsuvina T, 'Artificial Intelligence Technologies in the Judiciary: European Standards and Ukrainian Practice' (2025) 139(2) Foreign Trade: Economics, Finance, Law 4. doi:10.31617/3.2025(139)03

24. Wrzaszcz P, 'E-justice in Poland – Polish Experiences' (2023) 16(1) Teka Komisji Prawniczej PAN Oddział w Lublinie 381. doi:10.32084/tkp.5288

AUTHORS INFORMATION

Oleh Syniehubov
D.Sc., Kharkiv Regional Military Administration
sinegubov.oleg@gmail.com
https://orcid.org/0000-0001-6362-3115
Co-author, responsible for conceptualization, methodology, project administration, supervision, validation, and writing – original draft.

Oksana Bortnik
PhD, Department of Civil, Labour and Commercial Law, Kharkiv National University of Internal Affairs
bortnikoksana1980@gmail.com
https://orcid.org/0000-0001-7816-0387
Co-author, responsible for investigation, formal analysis, and writing – review & editing.

Olena Chernenko*
PhD, Department of Public Management, Administration and Law, National University "Yuri Kondratyuk Poltava Polytechnic"
Department of Law, Private Higher Educational Institution "European University"
olena.chernenko22@gmail.com
https://orcid.org/0000-0003-4178-5417
Corresponding author, responsible for investigation, data curation, and writing – review & editing.

Competing interests: No competing interests were disclosed.

Disclaimer: The authors declare that their opinion and views expressed in this manuscript are free of any impact of any organizations.

RIGHTS AND PERMISSIONS

EDITORS

Managing editor – Dr. Olena Terekh. **English Editor** – Julie Bold.
Ukrainian language Editor – Mag. Lilia Hartman.

ABOUT THIS ARTICLE

Cite this article
Syniehubov O, Bortnik O and Chernenko O, 'European Approaches to Digital Justice in Central and Eastern Europe and the Baltic States, with Perspectives for Ukraine' (2025) 8(Spec) Access to Justice in Eastern Europe 32-73 <https://doi.org/10.33327/AJEE-18-8.S-c000162>

DOI: https://doi.org/10.33327/AJEE-18-8.S-c000162

***Keywords*:** *digital justice, e-justice, civil procedure, court decision, protection of rights, human-centred digital transformation, digital principles, digital rights, European Union, Ukraine, Central and Eastern Europe, Baltic States, AI.*

DETAILS FOR PUBLICATION

Date of submission: 07 Nov 2025
Date of acceptance: 13 Dec 2025
Date of Publication: 30 Dec 2025
Whether the manuscript was fast tracked? – Yes
Number of reviewer report submitted in first round: 2 reports + guest editor's editor
Number of revision rounds: 2 rounds with minor revisions

Technical tools were used in the editorial process
Plagiarism checks - Turnitin from iThenticate https://www.turnitin.com/products/ithenticate/
Scholastica for Peer Review https://scholasticahq.com/law-reviews

AI DISCLOSURE STATEMENT

The authors confirm that no artificial intelligence tools or services were used at any stage of writing, translating, editing, or analyzing content for this manuscript.

АНОТАЦІЯ УКРАЇНСЬКОЮ МОВОЮ

Дослідницька стаття

ЄВРОПЕЙСЬКІ ПІДХОДИ ДО ЦИФРОВОГО ПРАВОСУДДЯ В КРАЇНАХ ЦЕНТРАЛЬНОЇ ТА СХІДНОЇ ЄВРОПИ ТА БАЛТІЙСЬКОГО РЕГІОНУ З ОГЛЯДУ НА ПЕРСПЕКТИВИ ДЛЯ УКРАЇНИ

Олег Синєгубов, Оксана Бортнік та Олена Черненко*

АНОТАЦІЯ

Вступ: *У статті досліджуються європейські підходи до цифрового правосуддя з акцентом на тому, як наднаціональні регуляторні моделі – від принципів цифрових прав до операційних інструментів е-правосуддя – формують національні практики в країнах Центральної та Східної Європи та Балтійського регіону. У дослідженні цифрове правосуддя концептуалізується як багатовимірне явище, що поєднує технологічні інструменти, інституційний дизайн, управління даними та людиноцентричні цінності. Окрема увага приділена тому, яким чином ці європейські напрацювання можуть сприяти реформуванню сектору правосуддя України в умовах наближення її законодавства до acquis ЄС.*

Методи: *Дослідження ґрунтується на доктринальному правовому аналізі у поєднанні з компаративним підходом. Системно розглядаються нормативні акти ЄС, інструменти СЕРЕЙ, моніторингова архітектура «Цифрового десятиліття», а також законодавство з цивільного процесу та платформи е-правосуддя в одинадцяти державах-членах ЄС. Емпіричну базу доповнюють дані EU Justice Scoreboard 2024–2025 та Digital Decade Country Reports. Такий методологічний підхід дозволяє виявити спільні закономірності, відмінності та моделі імплементації, що формують основу для оцінки траєкторії цифрового правосуддя в Україні.*

Результати та висновки: *Дослідження показує, що Європейський Союз сформував цілісну, ціннісно орієнтовану архітектуру цифрового правосуддя, яка поєднує юридично обов'язкові стандарти, інтероперабельні технологічні рішення та принципи інклюзивності, прозорості й людського контролю – особливо у контексті високоризикових систем ШІ. Найбільш інтегровану та технологічно розвинену модель демонструють Балтійські держави, тоді як країни Центральної та Східної Європи характеризуються поступовими або фрагментованими підходами. Системна інтеграція – обов'язкове електронне подання, уніфіковані інфраструктури обміну даними та машиночитані судові дані – корелює з вищими показниками за індикаторами ЄС. Натомість фрагментовані або паралельні системи обмежують доступність, інтероперабельність і можливості правосуддя, заснованого на даних. Україна демонструє суттєвий поступ через ЄСІТС, автоматизоване управління справами та електронний документообіг; однак зберігаються суттєві прогалини у сфері інтероперабельності, машиночитаності, людиноцентричного дизайну та врядування ШІ. Структурна фрагментація та відсутність комплексної стратегії цифрового правосуддя стримують наближення до стандартів ЄС. На основі компаративного аналізу визначено пріоритети для України: повноцінна імплементація машиночитаних форматів (XML/JSON та ECLI), консолідація розрізнених підсистем в єдину екосистему, запровадження обов'язкових цифрових процедур для професійних учасників процесу, а також розроблення правозахисної моделі врядування ШІ, узгодженої з Європейським актом про ШІ. Ключовим завданням є забезпечення того, щоб цифровізація розширювала – а не обмежувала – доступ до правосуддя, вимагаючи збалансування технологічних інновацій із процесуальними гарантіями, інституційною стійкістю та інклюзією користувачів.*

Ключові слова: *цифрове правосуддя, е-правосуддя, цивільне судочинство, рішення суду, захист прав, людиноцентрична цифрова трансформація, цифрові принципи, цифрові права, Європейський Союз, Україна, Центральна та Східна Європа, Балтійські держави, ШІ.*

Access to Justice in Eastern Europe
ISSN 2663-0575 (Print)
ISSN 2663-0583 (Online)
http://ajee-journal.com

Research Article

THE NEW EU PRODUCT LIABILITY DIRECTIVE: DOCTRINAL ANALYSIS

Deimantė Rimkutė

ABSTRACT

Background*: The question of who bears liability when an AI system causes harm has long been debated in the field of AI and law. When the EU proposed the Product Liability Directive and the AI Liability Directive in 2022, many expected the two instruments to jointly determine how such cases would be handled. This picture changed once the AI Liability Directive was withdrawn in February 2025 and the revised Product Liability Directive was adopted in October 2024. The result is that the EU now provides harmonised liability rules only for situations in which AI products injure consumers or other natural persons. With this shift, current discussions on AI liability can either examine the consequences of withdrawing the AI Liability Directive or look closely at how the updated Product Liability Directive (EU) 2024/2853 ("Directive") allocates responsibility. This article takes the latter approach. For the consumer and natural-person contexts to which the Directive applies, it argues that the debate should turn to analysing the specific liability structure the Directive sets out.*

DOI:
https://doi.org/10.33327/AJEE-18-8.S-c000160

Date of submission: 31 Oct 2025
Date of acceptance: 05 Dec 2025
Online First Publication: 19 Dec 2025
Last Published: 30 Dec 2025

Disclaimer:
The author declares that their opinion and views expressed in this manuscript are free of any impact of any organizations.

Methods: *The revised Directive is not an easy instrument to interpret, given its distinctive terminology, high level of detail, and extensive lex specialis provisions when compared to the general rule of fault-based liability. Although it replaces the 1985 Directive–and thus builds on an existing body of scholarship–the 2024 revision introduces changes that necessitate renewed analysis. The most suitable method for understanding this revised liability regime, clarifying its ambiguities, novelties, and problematic aspects, is the doctrinal legal approach. Accordingly, this article employs doctrinal analysis to examine and systematise the revised Directive.*

Results and conclusions: *The article organises the Directive's provisions into four categories: first, the scope of application, defining when product liability applies; second, the elements of liability, specifying what the claimant must prove; third, the defences, indicating the exemptions from liability on which the defendant may rely; and finally, the procedural rules, governing disclosure of evidence, relevant to both parties, and the conditions under which the burden of proof, resting on the claimant, may shift to address evidentiary challenges.*

On the basis of this fourth feature—procedural rules that ease the claimant's evidentiary burden—the article argues that the Directive alters how EU product liability should be conceptually defined. Under the 1985 Directive, the framework rested on two conceptual axes defining strict product liability: first, replacing fault with product defectiveness; and second, limiting defences, both of which made product liability "strict". Since the new procedural rules further strengthen the claimant's position, the article concludes that the revised Directive adds a third axis to the concept of strict product liability, thereby making the regime even stricter through burden-alleviation rules.

1 INTRODUCTION

If an AI system causes damage, who is legally responsible for compensating the victim? This has been one of the central questions in contemporary debates in the field commonly referred to as "AI and law".[1] Since October 2024, however, the answer has become clearer—at least with respect to consumers and other natural persons. That date marks the adoption of the revised Product Liability Directive (EU) 2024/2853 ("Directive") .[2] This Directive now covers situations where damage is caused by AI,[3] meaning the debate

1 The author has explored this issue in a blog post, see: Deimante Rimkute, 'AI Liability After the AILD Withdrawal: Why EU Law Still Matters?' (*Oxford Business Law Blog*, 1 April 2025) <https://blogs.law.ox.ac.uk/oblb/blog-post/2025/04/ai-liability-after-aild-withdrawal-why-eu-law-still-matters> accessed 30 May 2025.

2 Full title: Directive (EU) 2024/2853 of the European Parliament and of the Council of 23 October 2024 On Liability for Defective Products and Repealing Council Directive 85/374/EEC [2024] OJ L 2853/1.

3 See information, for example, in 'Liability for Defective Products' (*European Commission: Internal Market, Industry, Entrepreneurship and SMEs*, 8 December 2024) <https://single-market-economy.ec.europa.eu/single-market/goods/free-movement-sectors/liability-defective-products_en> accessed 30 May 2025.

should shift from identifying who bears responsibility to investigating how that responsibility is structured and applied in practice.

This question, however, is not straightforward. The Directive is a complex instrument, characterised by distinctive terminology, a high level of regulatory detail, and extensive *lex specialis* provisions. Although it replaces the 1985 Directive–and thus its commentary rests on an existing body of scholarship—the 2024 revision introduces changes that necessitate additional doctrinal work. This is where the present article makes its contribution: it offers a doctrinal analysis[4] of the revised Directive, aiming to provide a coherent and systematic understanding of its structure, key features, and novelties.

This article pursues this aim by grouping the Directive's provisions into four categories. The first concerns the starting point in the Directive's applicability—the scope of application, which defines whether disputes fall under the product liability regime (Section 2). The second category addresses matters forming the evidentiary burden of the claimant—namely, the elements of liability that must be proven to obtain compensation (Section 3). The third examines the defendant's escape routes from liability. That is, the exemptions from liability, invoked by the defendant to avoid responsibility (Section 4). Finally, the fourth category concerns the procedural rules, proposing remedies to evidentiary difficulties (Section 5).

By organising the Directive's provisions in this manner, the article also advances a broader conclusion about how product liability should be defined as a legal concept (Section 1). It argues that the 2024 Directive requires a reconsideration of this concept: product liability should no longer be understood solely as faultless liability—defined by product defectiveness and by narrowly circumscribed defences—but also as a regime that includes a third, procedural element, namely the alleviation of the claimant's burden of proof. The discussion begins with this revised definition of the concept of product liability.

2 DEFINING PRODUCT LIABILITY

Product liability is commonly defined as "no-fault" liability for a particular factual situation—damage caused by defective products. This standard definition, however, is not particularly helpful, as it merely explains what this type of liability is not (i.e., not fault-based) or limits itself to describing the circumstances in which it applies. The aim of this section, therefore, is to propose a more precise conception of product liability in 2024.

The starting point of this inquiry is the observation that, in the EU—unlike in the common law tradition—product liability is a regime established by legislation. Any conception of it must therefore begin with what the law itself provides, rather than how courts have defined

4 More about the methodology here: Jan M Smits, 'What is Legal Doctrine?: On the Aims and Methods of Legal-Dogmatic Research' in Rob van Gestel, Hans-W Micklitz and Edward L Rubin (eds), *Rethinking Legal Scholarship: A Transatlantic Dialogue* (CUP 2017) 210. doi:10.2139/ssrn.2644088.

it over time. In the EU, the law itself shows that the development of product liability has unfolded in two distinct phases. The first phase refers to EU product liability between 1985 and 2024.[5] This phase began in 1985, when the Product Liability Directive was adopted in response to a series of major industrial disasters that exposed the inadequacy of fault-based liability in addressing such cases.[6] The second phase begins in 2024, when the 1985 Directive was replaced by the 2024 Directive in response to different concerns—namely, growing criticism that the 1985 Directive was no longer adequate to address the challenges posed by new types of goods, particularly those involving digital technologies.[7] This brief overview matters for our task of defining the concept of product liability for one simple reason. If such a definition of the concept must be grounded in the law, and the law has developed in two distinct phases, then we must also consider whether there are two corresponding conceptions of product liability. In other words, while the classic understanding of product liability derived from the 1985 Directive, we should not assume that this definition still fully applies today. To assess whether the concept has changed, the understanding of product liability implied in the 1985 Directive must be clarified.

Under the 1985 Directive, the features that defined product liability as a *lex specialis*, strict, or, in other words, a victim-friendly liability regime were twofold. The first was that, under this regime, fault was replaced by another, so-called "strict" element—product defectiveness. This framed the regime as an *obligation de résultat*: the manufacturer had to ensure that the product was safe and did not cause harm, rather than merely exercising due care to make it safe (*obligation de moyens*).[8] This made product liability stricter because it was easier to prove the strict element of product defectiveness than to establish fault. That assessment was outcome-based—the product had to be made in a way that avoided defectiveness. The second axis was that the product liability regime limited and predetermined the defences available to the defendant, so that liability could be avoided only in clearly defined and exceptional circumstances,[9] for example, by showing that the

5 We do not consider period before 1985 because then product liability claims were largely governed by fault-based rules applied to a specific factual matrix – damage caused by a product. Read about it here: Duncan Fairgrieve and others, 'Product Liability Directive' in Piotr Machnikowski (ed), *European Product Liability: An Analysis of the State of the Art in the Era of New Technologies* (Intersentia 2016) 17.

6 ibid 20.

7 Read about the debate here: Sebastian Lohsse and others, 'Liability for Artificial Intelligence' in Sebastian Lohsse, Reiner Schulze and Dirk Staudenmayer (eds), *Liability for Artificial Intelligence and the Internet of Things* (Nomos 2019) 9. doi:10.5771/9783845294797-9; K Alheit, 'The Applicability of the EU Product Liability Directive to Software' (2001) 34 Comparative and International Law Journal of Southern Africa 188; Daily Wuyts, 'The Product Liability Directive – More than Two Decades of Defective Products in Europe' (2014) 5(1) Journal of European Tort Law 1. doi:10.1515/jetl-2014-0001; Gerhard Wagner, 'Software as a Product' in Sebastian Lohsse, Reiner Schulze and Dirk Staudenmayer (eds), *Smart Products* (Nomos 2022) 164. doi:10.5771/9783748929772-157.

8 Thomas Verheyen, 'Modern Theories of Product Warnings and European Product Liability Law' (2019) 15(3) Utrecht Law Review 44. doi:10.36633/ulr.541.

9 Cees van Dam, *European Tort Law* (3rd edn, OUP 2013) para 1003-3, 301.

defect could not have been detected given the state of scientific and technical knowledge at the time. This approach contrasted with fault-based liability, where the range of possible defences was much broader and less favourable to victims.

Turning now to the concept of product liability under the 2024 Directive, it becomes apparent that the regime maintains the two original axes. Product liability can still be defined as a form of no-fault, or strict, liability in which (1) fault is replaced by another element (product defectiveness) and (2) the available defences are narrowly limited. However, the initial doubt as to whether the concept remains truly the same reveals that the Directive introduces an additional structural feature. This third axis is the use of presumptions—mechanisms capable of shifting the burden of proof in establishing defectiveness and causation from the claimant to the defendant (discussed in Section 5).

The article argues that this new feature should be recognised as a third axis that further defines product liability alongside the two already described. This third feature renders product liability under the 2024 Directive even more "strict" and more favourable to claimants. The idea that presumptions make a liability regime stricter is not entirely new. Similar arguments have been made in discussions of fault-based liability, where regimes in which fault is presumed were described as *de jure* fault-based but *de facto* strict (or semi-strict), since their practical outcomes closely resemble those of strict liability.[10] Accordingly, to understand what constitutes "strict product liability," we must consider all the structural elements that contribute to its strictness.

3 SCOPE OF APPLICATION

The scope of application is always the starting point, because when damage occurs, the first question is which liability regime applies—product liability or another. At first glance, the question of the scope of application may seem relatively innocuous and academic attention often focuses on matters such as the elements of liability or defences. In practice, however, the scope of application is anything but innocuous. Because the Directive establishes a regime that is generally more favourable to claimants, defendants have strong incentives to argue that a given case falls outside its scope, while claimants will argue the opposite.

These disputes usually focus on three key concepts. The first is what qualifies as a product, since the Directive only applies if the damage was caused by something that meets the Directive's definition of a "product." The second concerns who can be held liable, as a claim cannot be brought against just anyone involved with the product. The third concerns who is entitled to claim damages, because the Directive does not confer standing to all persons

10 See the overview of the debate on the use of presumptions to establish fault in this article: Francesco Parisi and Giampaolo Frezza, 'Burdens of Proof in Establishing Negligence: A Comparative Law and Economics Analysis' (2023) 9 Italian Law Journal 77.

without distinction. The following paragraphs examine each of these concepts in more detail, beginning with the definition of a "product."

Product. The provision, which defines which goods fall under the Directive, is Article 4(1). It defines products as "all movables, even if integrated into, or interconnected with, another movable or an immovable; it [also] includes electricity, digital manufacturing files, raw materials, and software." This definition entails several notable changes. The first notable change is that the Directive is no longer limited to tangible goods. It now explicitly includes certain non-tangible items—digital manufacturing files and software (including AI).[11] This expansion, however, is not without limits. While software is generally covered, free and open-source software[12] is not (Article 2(2), Recital 14). The same applies to digital files: when such files are manufacturing files—for example, those used in advanced production processes such as CAD, CAM, or IoT-based systems, they are included, whereas generic digital files such as text documents, images, or videos are not.[13]

Another notable change in the Directive definition of a product is that it now also covers raw materials.[14] More revolutionary, it extends its scope to certain product-related services,[15] but only where two conditions are met: (1) the service is embedded in, or essential to, the product's functionality (Article 4(3)); and (2) it has a direct impact on the product's safety.[16] This means that services such as traffic data for vehicle navigation, health-monitoring functionalities, smart-home safety systems, and software updates, upgrades, or machine-

11 Directive (EU) 2024/2853 (n 2) recital 13. Although the core text does not define software, the Recitals of the Directive offer a non-exhaustive list, including operating systems, firmware, computer programs, applications, and AI systems. They also clarify that software is covered regardless of mode of supply, whether it is a standalone product, integrated component, stored on a device, accessed via cloud or network, or provided as software-as-a-service (SaaS)..

12 Free and open source software is typically governed by licenses that permit anyone to run, copy, distribute, study, and improve it; and must not be provided during commercial activity, for example, "in exchange for a price, or for personal data used other than exclusively for improving the security, compatibility, or interoperability of the software." See, Directive (EU) 2024/2853 (n 2) recital 14.

13 ibid, recital 16.

14 Importantly, the revised definition now encompasses materials that are not manufactured but are extracted or harvested, including raw materials and primary agricultural products, such as gas, water, farm goods, fishery products, and game, which were previously excluded under the 1985 Directive. See, Council Directive 85/374/EEC of 25 July 1985 On the Approximation of the Laws, Regulations and Administrative Provisions of the Member States Concerning Liability for Defective Products [1985] OJ L 210/29, art 2. For the context see, for example, Duncan Fairgrieve and Richard Goldberg, *Product Liability* (3rd edn, OUP 2020) 221. doi:10.1093/oso/9780199679232.001.0001.

15 See, for example, Directorate-General for Communication, 'EU Adapts Product Liability Rules to Digital Age and Circular Economy' (*European Commission*, 9 December 2024) <https://commission.europa.eu/news/eu-adapts-product-liability-rules-digital-age-and-circular-economy-2024-12-09_en> accessed 7 November 2025.

16 See, Directive (EU) 2024/2853 (n 2) recital 17. Also, Gerhard Wagner, 'Next Generation EU Product Liability – For Digital and Other Products' (2024) 15(2) Journal of European Tort Law 185-6. doi:10.1515/jetl-2024-0011.

learning processes may now fall within the Directive's scope.[17] The Directive, therefore, applies "well beyond the realm of both pure manufacturing and products *sensu stricto*".[18]

Caution is nevertheless required when applying the Directive to services. The inclusion of related services should not lead to the misunderstanding that the Directive applies to services in general. Pure information content—such as media files, e-books, source code, or other forms of data that merely constitute "information"— remains outside the Directive's scope.[19] In this regard, the clarification given by the CJEU in *Krone-Verlag* (Case C-65/20) outlives the Directive's revision: harm caused solely by incorrect information contained in a medium such as a newspaper does not make that medium a defective product under the PLD.[20]

Persons entitled to claim damage. A second gatekeeper from the application of the Directive concerns who is permitted to bring a product liability claim. The Directive does not grant standing to any person in general. It grants standing only to natural persons who have suffered harm as a result of a defective product (Article 5(1)).[21] This, however, raises a question: does the Directive apply in the same way to all natural persons? That would seem counterintuitive, as product liability has traditionally been understood as a consumer-oriented regime. To answer this, Article 5 must be read together with Article 6, which defines "damage." Article 6 makes it clear that certain categories of compensable harm remain limited to non-professional or not exclusively professional use. Accordingly, while the Directive extends protection to all natural persons, particularly in relation to damage to health, it provides a narrower protection for those acting in a professional capacity, where the damage concerns property or data. In that sense, the Directive formally protects all natural persons but, in practice, offers stronger protection to consumers.

Liable parties. The last gatekeeper from the application of the directive concerns liable parties. A claimant seeking compensation under the Directive cannot sue just anyone involved with the product. Only certain parties—those explicitly listed in the Directive—can be held liable. Importantly, this may not always be the company that actually manufactured the defective product. Liability can extend further down the supply chain to others who played a role in bringing the product to the market. To capture this broader scope, the Directive uses the term "economic operators." This includes not only the manufacturer (of the product or one of its components) but also the provider of a related service, the authorised representative, the importer, the fulfilment service provider, and, in some cases, the distributor (Article 4(15)).

17 Directive (EU) 2024/2853 (n 2) recitals 17, 19, 35.

18 Jan De Bruyne, Orian Dheu and Charlotte Ducuing, 'The European Commission's Approach to Extra-Contractual Liability and AI: An Evaluation of the AI Liability Directive and the Revised Product Liability Directive' (2023) 51 *Computer Law & Security Review* 105894. doi:10.1016/j.clsr.2023.105894.

19 See, Directive (EU) 2024/2853 (n 2) recital 13.

20 Case C-65/20 *VI v KRONE-Verlag Gesellschaft mbH & Co KG* [2021] ECLI:EU:C:2021:471, para 42.

21 It shall be noted that the revised Directive also includes persons who have inherited this right or to whom it has been subrogated, as well as persons acting on behalf of one or more injured persons. See, Directive (EU) 2024/2853 (n 2) art 5(2).

In assigning the responsible party, the Directive follows a "chain of liability" principle: there is a set order in which potential defendants are approached. If the first entity in the chain (e.g., the manufacturer) cannot be held liable, the claim moves to the next party (e.g., importer). In total, there are four tiers in this chain of liability. The first tier consists of manufacturers, a term that the Directive defines broadly. It includes two groups. First are the "real manufacturers"—those who actually produce the product, including when they make it for their own use[22] (Article 4(10)(a), (c)). Second are "manufacturers by labelling"—parties that have a product made by someone else but place it on the market under their own name, trademark, or another identifying sign (Article 4(10)(b)). Although these two categories involve different levels of control over design and production, the Directive does not require the claimant to track down the "real manufacturer" before suing a "manufacturer by labelling." This approach was confirmed by the CJEU in *Fennia* (Case C-264/21) and remains relevant even after the Directive's revision.[23]

When the manufacturer is not established or cannot be identified within the EU (Article 8(3)), the Directive allows the claimant to pursue other economic operators in the supply chain. Before listing which operators may be held liable, it is necessary to explain why this expansion exists and what purpose it serves. The underlying rationale reflects the Directive's victim-protection logic: its objective is to ensure that those harmed by defective products have a realistic chance of obtaining compensation when the manufacturer (the primary responsible party) cannot be reached.

The criteria that bound these other liability parties—namely, importers, authorised representatives, fulfilment service providers, and distributors—is that they are not random third parties. In factual terms, each operator can contribute to the process by which unsafe products reach consumers through import, handling, storage, or sale. At the same time, these economic operators profit from the activity that results in the unsafe product being placed on the market.[24] If they take part in an activity that generates risk and derive economic benefit from it,[25] it is reasonable that they also bear part of that risk and be held liable when the manufacturer is unavailable, and the product causes harm.

22 It appears that with this addition to include into manufacturers notion also the manufactures that manufacture for their own use codifies the CJEU's judgment in *Veedfald* (Case C-203/99). In that case, the Court held that a service provider may be liable under the Directive where it uses a self-manufactured product in the course of providing a service and the defect lies in that product, rather than in the service itself. See, Case C-203/99 *Veedfald v Århus Amtskommune* [2001] ECR I-3569, ECLI:EU:C:2001:258, para 12.

23 Case C-264/21 *Keskinäinen Vakuutusyhtiö Fennia v Koninklijke Philips NV* [2022] ECLI:EU:C:2022:536, para 35.

24 Marco Cappelletti, *Justifying Strict Liability: A Comparative Analysis in Legal Reasoning* (OUP 2022) 73-4. doi:10.1093/oso/9780192859860.001.0001.

25 Read about risk creation as justification for strict liability in Guido Calabresi, *The Costs of Accidents: A Legal and Economic Analysis* (Yale UP 1970) 50-4.

However, not all operators contribute equally to this process. The Directive recognises this, giving some operators higher priority than others. This further hierarchy appears to rest on two main factors: the operator's role in placing the product on the market or making it available to users, and their degree of proximity to the manufacturer. Based on these criteria, the second possible defendants, if the manufacturer is not established in the EU, are the importers and authorised representatives (Article 8(1)(c)). These actors have the closest connection to both the manufacturer and the product's entry into the EU market, which explains why they are second. The importer places the product on the EU market and bears compliance responsibility (Article 4(12)), while the authorised representative acts as the manufacturer's proxy, carrying out specific duties related to product compliance and communication with authorities under a defined mandate.[26]

The third-tier operators are fulfilment service providers[27]—those who handle key logistics tasks such as warehousing, packaging, shipping products (Article 8(1)(c)). The claimant may turn to them only if no importer or authorised representative is established within the Union. Their inclusion recognises their practical role in enabling products, particularly from outside the EU, to reach consumers, bypassing standard product placement procedures designed to ensure safety.[28] This is especially relevant in the context of cross-border e-commerce, where fulfilment service providers often act as *de facto* importers.[29] At the bottom of the hierarchy are the fourth-tier operators—distributors and online platforms (Article 8(3)-(4)).[30] They serve as "last-resort" defendants, liable only

26 See, for example, Netherlands Enterprise Agency and Netherlands Food and Consumer Product Safety Authority, 'Product Safety and the Role of the Authorised Representative' (*Business.gov.nl*, 2024) <https://business.gov.nl/regulation/product-safety-and-role-of-authorised-representative/> accessed 7 June 2025. Importantly, authorised representatives can be assigned additional administrative tasks (e.g., drafting the EU declaration of conformity or affixing CE marking) may be delegated by the manufacturer when permitted by law and the mandate.

27 Directive (EU) 2024/2853 (n 2) art 4(13). The Directive defines fulfilment service provider as "any natural or legal person offering, in the course of a commercial activity, at least two of the following services: warehousing, packaging, addressing and dispatching of a product, without having ownership of that product". This definition excludes postal services (art. 2(1) of Directive 97/67/EC), parcel delivery services (art. 2(2) of Regulation (EU) 2018/644), and other postal or freight transport services.

28 Read about it, for example, here: 'What Does It Mean "Placing a Product on the Union Market"?' (*European Commission: Energy Efficient Products,* 2024) <https://energy-efficient-products.ec.europa.eu/faqs-0/what-does-it-mean-placing-product-union-market_en> accessed 7 June 2025.

29 See, Directive (EU) 2024/2853 (n 2) recital 37.

30 It is important to note, however, that the inclusion of online platforms in this tier is conditional: they can be held liable only if the criteria in Article 6(3) of the Digital Services Act are met – that is, where the platform actively presents the product or otherwise enables the specific transaction (Article 8(4)). See, Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act) [2022] OJ L 277/1. On the other hand, if online platforms function as a manufacturer, importer, authorized representative, fulfilment service provider, or distributor, then they take on the same liabilities as those operators, i.e. Article 8(4) is not applied. See, Directive (EU) 2024/2853 (n 2) recital 38.

when the claimant cannot identify a higher-tier operator. Their inclusion can be justified by the "soft power" they have, i.e. they decide which products to distribute or host on their platforms and can therefore influence whether unsafe products reach consumers while benefiting from this activity. However, since they do not control a product's initial entry into the market, they are placed at the bottom.

4 ELEMENTS OF LIABILITY

To obtain compensation, the claimant must prove that the damage was caused by a defective product. In other words, defectiveness and damage must be shown to stem from the same event, with the damage occurring as a result of the defect. This assessment is structured around three elements: product defectiveness, damage, and causation. Under the general rule, it is the claimant's responsibility to establish all three. This section, however, will address only two elements: product defectiveness and damage. Causation is regulated by the Directive only to the extent that it is presumed and will therefore be examined separately in Section 5.

Damage. The starting point of any delict claim is actual damage—without it, there is no basis for compensation. The claimant must therefore show that a product defect caused real harm, both factually and in monetary terms. Damages are generally classified as economic (quantifiable financial losses such as medical or funeral expenses, property damage, and lost income)[31] and non-economic (intangible harm such as pain and suffering, emotional distress, loss of enjoyment of life, or loss of companionship).[32] However, many legal systems restrict recovery for certain types of harm—typically pure economic loss or emotional distress.

The Product Liability Directive adopts a similar approach by limiting compensable damage to three categories: (1) damage to health, including medically recognised psychological harm; (2) damage to property; and (3) damage to data. This means that infringements of personality rights, privacy breaches, acts of discrimination, and pure economic loss (e.g. loss of salary) fall outside its scope.[33] The Directive then narrows these categories further through specific exclusions. It does not cover damage to the defective product itself; damage to another product caused by a defective component where both are made or controlled by the same manufacturer; property used exclusively for professional purposes; or destruction or corruption of data used for professional purposes (Article 6(1)(b)–(c)).

On the other hand, while Directive does not cover every type of damage, that does not stop Member States from allowing those excluded types of damage to be compensated under

31 David A Fischer and others, *Products Liability: Cases and Materials* (West Academic Publishing 2022) 673.
32 ibid
33 See, Directive (EU) 2024/2853 (n 2) recital 24.

their own national law.[34] They can do this in two ways: through another national liability regime, such as fault-based liability, or by extending the product liability regime to additional types of damage. If they choose the latter option, they may replicate the same rules set out in the Directive. But in that case, the legal basis is national law, not EU law. The Member State is essentially creating a parallel national regime inspired by the Directive rather than implementing the Directive itself. CJEU case law confirms this approach. In *Moteurs Leroy Somer v Dalkia France* (Case C-285/08), the CJEU held that damage to exclusively professional property falls outside the scope of the PLD, but Member States are free to adopt equivalent national regimes.[35]

Product defectiveness. The concept of "defect" is considered to be the core concept, upon which product liability turns,[36] and is widely regarded as the most complex element of the product liability.[37] The Directive addresses it through two main tests: the classic "consumer expectations test", and a second – newly introduced – test that can be called "safety required by law test" (Article 7(1)). To assist in the process, the Directive also introduces a list of circumstances that must or must not be (solely) considered when assessing whether a product is defective, outlined in Article 7(2) and (3). The main uncertainty, however, concerns the new test – introduced late in the trilogue negotiations – whose meaning sparks the debate.[38] This article, therefore, begins with that second test.

At first glance, the "safety required by law" test might suggest that any failure to meet legal safety standards automatically makes a product defective. However, when the Directive is read in full, this conclusion proves too broad. References to what is "required", whether by law or by other relevant requirements, appear in different provisions and lead to different legal consequences. For instance, Article 10(2)(b) establishes a presumption of defectiveness where a product breaches mandatory Union or national safety rules specifically aimed at preventing the kind of harm that occurred. By contrast, Article 7(2)(f) treats "relevant product safety requirements, including safety-relevant cybersecurity requirements" as merely one factor among several in the overall assessment of defectiveness. The question, then, is how these provisions relate to one another. This article takes the view that not every regulatory breach renders a product unsafe, and therefore not every breach renders it defective. As Table 1 shows, the legal consequences depend on the nature of the breach and its impact on the product's safety. If the breached rule directly renders the product unsafe, Article 7(1) applies, and the product is defective. If the breach raises a legitimate safety concern but does not in itself demonstrate unsafety,

34 ibid, recitals 23, 24.

35 Case C-285/08 *Moteurs Leroy Somer v Dalkia France and Ace Europe* [2009] ECR I-4733, ECLI:EU:C:2009:351, paras 28, 31.

36 Fischer and others (n 31) 115.

37 Fairgrieve and Goldberg (n 14) 324.

38 For example, considering remarks by, Piotr Machnikowski, 'Defectiveness' (The New Product Liability Directive: Doctrinal, Comparative and Interdisciplinary Approaches: Conference, Maastricht University, 17-18 October 2024).

Article 10(2)(b) creates a rebuttable presumption of defectiveness. If the relevant requirement is not established in law, or is not breached but still relevant in context, it serves only as a factor in the assessment under Article 7(2)(f).

Table 1.

Article / Concept	Purpose	Meaning
Art. 7(1)	Establishing an actual defect	A breach of a legal rule establishes defectiveness only if (1) the rule is established in law (EU / national), and (2) the breach actually makes the product unsafe.
Art. 10(2)(b)	Presumption of defect (evidentiary facilitation)	A presumption of defectiveness arises if (1) mandatory product safety requirements laid down in law (EU / national) are breached, and (2) those requirements are intended to protect against the relevant risk of damage. This presumption does not require proof that the breach made the product unsafe, but if it did, then Art. 7(1) applies.
Art. 7(2)(f)	Assessment factor	Product safety requirements shall be assessed when investigating defectiveness. They do not need to be established in law (national / EU) or breached. However, if they are established in law (national / EU) and breached, then Art. 7(1) or Art. 10(2)(b) may become applicable.

Source: Author

Next, the classic consumer expectations test is broader and more open-ended than the "safety required by law" test. While the factors listed in Article 7(2) are technically relevant to both, in practice, they offer greater guidance when applying the consumer expectations test. A useful starting point in discussing consumer expectations is to identify the type of defect that caused the damage. Doctrine distinguishes three categories: manufacturing defects (the design is sound, but production errors make the product unsafe),[39] design defects (the design itself fails to provide adequate safety),[40] and information defects (insufficient instructions or warnings).[41] These categories shape what a reasonable consumer may expect. For manufacturing defects, consumers can *prima facie* expect almost full conformity with the intended safe design; any safety-relevant

39 Fischer and others (n 31) 115.

40 ibid

41 Sanne B Pape, *Warnings and Product Liability: Lessons Learned from Cognitive Psychology and Ergonomics* (Eleven International Publishing 2012) 277, also Fairgrieve and Goldberg (n 14) 355.

deviation will generally fail the consumer expectations test.[42] A reasonable consumer expects a product to be correctly manufactured, not that the manufacturer might make preventable errors in production. By contrast, for design and information defects, establishing defectiveness is not as straightforward.[43] Expectations depend more heavily on the surrounding context, in which the circumstances listed in Article 7(2) become relevant. A practical approach is therefore: first ask whether the issue is a manufacturing defect; if not, proceed to the broader Article 7(2) assessment.

Article 7(2) sets out a broad – and importantly, non-exhaustive – list of circumstances that may be considered when assessing defectiveness. Several points emerge from analysing these factors. First, some of them can be read as effectively implying (primary) duties for the producing process, and their breach may strongly support a finding of defectiveness. This is evident, for example, in the obligation to consider the specific needs of the intended users (Article 7(2)(h))[44] or the effects the product may have when used together with other products (Article 7(2)(d)). A failure to consider these factors may constitute a strong argument for defectiveness. Second, if these circumstances are understood to imply duties of production, it becomes apparent that the Directive formulates these duties at different levels of strictness. In some cases, the wording almost suggests that failing to satisfy a given circumstance effectively implies defectiveness, for instance, where a product fails to fulfil its safety function in situations where the very purpose of it is to prevent harm (Article 7(2)(i)). In others, the Directive simply calls for consideration of reasonably foreseeable use (Article 7(2)(b)), leaving room for a more context-sensitive assessment.

Third, some circumstances do not necessarily impose duties on operators but instead serve a contextual function, guiding the time or object of the assessment. For example, Article 7(2)(a) concerns the object of the assessment, namely how the product is presented, its labelling, design, technical features, composition, packaging, and instructions for assembly, installation, use, or maintenance. Article 7(2)(e), meanwhile, directs the assessment of defectiveness by "locking" it to the moment when the product was placed on the market or left the manufacturer's control, on the logic that the manufacturer can only impact safety while it still has control over the product. At the same time, it communicates when this moment should be "unlocked", i.e., when the product is no longer under the manufacturer's ongoing control, as is often the case with

42 Van Dam (n 9) 427-8.

43 ibid

44 For this circumstance the relevant Directive (EU) 2024/2853 (n 2) recital 30. It highlights that fair apportionment of risks, especially when a product poses "high risks of causing damage," gives rise to "particularly high safety expectations." This principle has also been affirmed by the CJEU, which, in a case concerning a heart stimulator, recognized that patients have particularly high expectations of the safety of medical devices. For the reference, see joined Cases C-503/13, C-504/13 *Boston Scientific Medizintechnik GmbH v AOK Sachsen-Anhalt – Die Gesundheitskasse and Betriebskrankenkasse RWE* [2015] ECLI:EU:C:2015:148, para 39.

digital products. This implies that the moment from which assessment of such products is not necessarily fixed to their placement on the market.

On the other hand, as already noted, the Directive not only sets out the circumstances to be considered when assessing defectiveness, it also identifies circumstances that cannot be relied upon (at least not on their own). Article 7(3) is a key provision here. It states that a product cannot be considered defective solely because a better or updated version has been, or is later, placed on the market. When viewed in the broader tradition of product liability, this provision might at first appear to restrict the approach taken in the US, in particular, the alternative design test, which is closely associated with the so-called risk-utility test.[45] Under this test, claimants can argue that a practical and cost-effective safer design[46] – often demonstrated by existing products,[47] could have prevented the harm.[48] If the safety benefits of that alternative design outweigh the costs of implementation, the product is considered defective.[49]

The fact that the EU rejects an unrestricted version of this test reflects a clear normative choice—prioritising consumer safety over economic efficiency. As Taschner, one of the Directive's original drafters, argued, the risk-utility test ultimately favours producers, whereas the European approach is that if a manufacturer cannot place a sufficiently safe product on the market, it should not place it on the market at all.[50] However, this does not mean that alternative designs or risk-utility reasoning must be entirely disregarded. A textual reading of Article 7(3) confirms that the alternative design test is not excluded from the defectiveness analysis; it is simply not decisive on its own. A safer or more advanced alternative design may still inform the overall assessment of defectiveness when considered alongside other relevant factors. In this sense, "alternative design" reasoning can operate as supporting evidence within the Article 7(2) framework.[51]

45 Fairgrieve and Goldberg (n 14) 336-9.

46 Examples include a commercial coffee urn that exploded due to the absence of a simple pressure-reducing valve; a lawn chair with a metal mechanism under the armrest that severed a user's finger, where a protective housing could have averted the injury; or an industrial machine with a sharp edge that caused harm, where the edge served no purpose and could have easily been smoothed for safety. See more, David G Owen, *Products Liability in a Nutshell* (10th edn, West Academic Press 2023) 225.

47 ibid

48 David G Owen, *Products Liability Law* (3rd edn, West Group 2015) 300.

49 ibid 303.

50 Hans Claudius Taschner, 'Product Liability: Basic Problems in a Comparative Law Perspective' in Duncan Fairgrieve (ed), *Product Liability in Comparative Perspective* (CUP 2005) 160. doi:10.1017/CBO9780511493850.011.

51 The indirect implication of the argument presented here is that, while alternative design cannot serve as the sole argument, other individual factors expressly listed in Article 7(2) of Directive (EU) 2024/2853 may, in principle, be sufficient on their own.

5 EXEMPTIONS FROM LIABILITY

As is typical in delict cases, even where all elements of liability are established, the defendant may still avoid liability by invoking an exemption, a circumstance which, if proven, releases the operator from liability. In the Directive, such exemptions are set out in Article 11 and, arguably, also in Article 13. Conceptually, these exemptions appear to be united by a common criterion: the operator is not liable where the risk that led to the damage lay outside their objectively reasonable sphere of control. In this sense, the exemptions implement a principle of fairness – liability under the PLD is strict but not absolute; the operator should not be held liable where it was not within their power to prevent the product from being defective.

To start with, exemptions of the Directive can be categorised into four broader categories. The first category covers situations where liability is exempted because the relevant operator did not place the product on the market (in the case of a manufacturer or importer, Article 11(1)(a)) or did not make it available on the market (in the case of a distributor, point (b)).[52] Accordingly, if the operator was not involved in introducing the product to the consumer and therefore did not contribute to the defective product reaching them, they should not be held responsible for the resulting damage.[53]

The second category concerns the absence of a defect at the time it was placed, put into service, or made available on the market, as provided in Article 11(1)(c). Accordingly, if the product had no defect when it entered circulation, the operator cannot be held liable. This exemption ensures that the operator is responsible only for risks within its sphere of control and cannot be held accountable for changes occurring after the product has left that sphere. However, precisely for this reason, an important limitation applies to digital products. Where the operator retains control, i.e. over software, related services, safety updates, or subsequent modifications, and the defect stems from those controlled elements, the operator cannot rely on this exemption (Article 11(2)).

The third category consists of compliance-based (Article 11(1)(d)) and development-risk (point (e)) defences. Under these, the defendant is not held liable where the product was made in accordance with the legal or state-of-the-art standards available at the relevant time. This reflects the idea that it would be unfair to penalise an operator for following the law[54] or to expect them to prevent what was objectively impossible to detect and fix. While the compliance-based defence is relatively clear, whether the product complied with

52 By analogy, this should be applied to online platform. See, of Directive (EU) 2024/2853 (n 2) recital 37, "provisions of this Directive relating to distributors should apply analogously to such online platforms."

53 This exemption is not applied to authorised representative because it only represents manufacturer, it does not directly place it on the market, so too fulfilment service provider, whose role is logistical, to ensure that the product reaches the consumer.

54 Geraint Howells (ed), *Law of Product Liability* (2nd edn, LexisNexis Butterworths 2007) 403.

requirements or not, the development risk defence is more open-ended in nature. In *Commission v United Kingdom* (Case C-300/95), the CJEU clarified that this defence requires invoking three elements: (1) the defect could not have been discovered given the highest level of scientific and technical knowledge available at the time; (2) that knowledge must have been objective; and (3) it must have been reasonably accessible,[55] as AG Tesauro pointed out—assessed in light of the actual opportunities for such information to circulate.[56]

The fourth category consists of situations where the damage is attributable to another person, not the defendant. The Directive expressly allows this defence in two cases. First, where the defendant is a component manufacturer, the defect in the final product is attributable to the design of that product or to the instructions given by its manufacturer, rather than to the component itself (Article 11(1)(f), referring to Article 8(1)(b)). Second, where the defendant is a modifier, the defect concerns a part of the product that was not affected by the modification (Article 11(1)(g), referring to Article 8(2)).

Conceptually, this fourth category can also be complemented by a third situation, mentioned in Article 13(2). This provision establishes contributory negligence, which ensures that the operator is not held fully or partly responsible where the damage was caused, or partially caused, by the injured party's fault. The Directive does not specify whether "fault" refers to ordinary or gross negligence. However, Recital 55 provides guidance through an example: a user failing to install essential safety updates, conduct which may not, at *prima facie*, amount to gross negligence unless the user was under a clear obligation to do so. Nevertheless, beyond the specific cases covered by Article 11 and Article 13(1), the Directive generally does not allow reliance on the acts or omissions of a third party to reduce or exclude liability (Article 13(1)), for example, where a hacker exploits a cybersecurity vulnerability and contributes to the damage.[57] In such cases, the operator remains fully liable to the injured person.

6 PROCEDURAL MATTERS

A key novelty of the Directive is that it goes beyond substantive rules to introduce procedural provisions. The rationale for this development lies in the problem of evidentiary deficiency—situations in which claimants are unable to substantiate their claim due to a lack of necessary evidence. Even under a strict liability regime, this problem persists, particularly in cases involving technically complex products such as pharmaceuticals or AI systems, where it can be difficult for the claimant to prove that the product was defective or that it caused the damage. To address this, the Directive introduces procedural mechanisms

55 Case C-300/95 *Commission of the European Communities v United Kingdom of Great Britain and Northern Ireland* [1997] ECR I-2649, ECLI:EU:C:1997:255, para 29.

56 ibid, Opinion of Advocate General Tesauro, para 24 AG.

57 See, Directive (EU) 2024/2853 (n 2) recital 55.

designed to assist claimants, specifically concerning the disclosure of evidence and the allocation of the burden of proof, which will now be examined in more detail.

Evidence disclosure. The evidentiary starting point in any dispute is access to the evidence. In AI-related product defect cases, for example, this may include logs, training data, or internal performance records, the kind of information that could reveal whether the product was defective. Yet the claimant, who bears the burden of proving defectiveness, typically has no access to such material. Vandenbussche describes this as a problem of evidentiary asymmetry: one party carries the burden of proof, while the other controls the information needed to meet it.[58] In evidence law, such imbalances are addressed through disclosure rules, which compel the party in control of the evidence to provide it. While national civil procedures already allow for this in principle, the Directive goes a step further by introducing product-liability-specific disclosure rules in Article 9. These rules aim to balance the interests of both sides: they require courts to protect trade secrets (Article 9(5)), but at the same time, they must also ensure that evidence is provided "in an easily accessible and easily understandable manner" (Article 9(6)), which is particularly important for the claimant.

Presumptions. Evidentiary difficulties, however, do not end with disclosure. Even where the court orders the defendant to provide evidence, the defendant may still refuse to comply. In such a situation, the claimant is left unable to prove defectiveness because she simply cannot access the necessary information. Vandenbussche refers to this as a problem of "evidentiary impossibility".[59] In such a case, Directive proposes a consistent with evidence law response – presumption of defectiveness (Article 10(2)(a)). This shifts the burden of proof to the defendant. In practice, this means that if the judge remains in doubt about the relevant facts at the end of the proceedings, the decision will be resolved against the party bearing the burden of proof—in this case, the defendant.

Disclosing evidence, on the other hand, is not a panacea in itself. Even with balanced disclosure rules, they may offer little help when the issue lies not in the unavailability of evidence, but in the limits of what can be known or demonstrated at all. Some facts are simply unprovable with a reasonable degree of certainty—for instance, when no existing scientific or technical knowledge can confirm or refute them. The law, however, does not remain indifferent to this problem of "evidentiary uncertainty".[60] A common response is to lower the standard of proof: instead of requiring the claimant to establish a fact with the usual level of certainty (i.e., a reasonable degree of certainty), courts may accept a less demanding threshold, such as proof based on probability.[61] This approach is also reflected

58 Wannes Vandenbussche, 'Dealing with Evidentiary Deficiency in Tort Law' (*SSRN,* 15 February 2019) doi:10.2139/ssrn.3335377 <https://ssrn.com/abstract=3335377> accessed 7 June 2025.

59 ibid 12.

60 Israel Gilead, Bernhard A Koch and Michael D Green (eds), *Proportional Liability: Analytical and Comparative Perspectives* (De Gruyter 2013) 328, para 11.

61 ibid

in the revised Directive. Under Article 10(4)(a-b), if the claimant faces excessive difficulties due to technical or scientific complexity, the standard of proof is reduced to what is "likely" when establishing defectiveness and/or causation. Accordingly, for instance, if an AI system makes a decision that results in physical harm and its reasoning cannot be fully reconstructed, the claimant may meet the burden of proof by showing that it is likely that a defect in the system contributed to the harmful outcome, even without identifying the precise algorithmic flaw.

Another group of presumptions in the Directive concerns neither evidentiary impossibility nor evidentiary uncertainty, but instead permits a presumption based on *prima facie* evidence—that is, evidence allowing the court to treat a fact as proven "at first sight," drawing on general experience that "if X occurs, then Y (the legally relevant fact) usually follows or shall be implied."[62] But before turning to the discussion of this group of presumptions, it is useful to recall the broader context in which they operate. While terminology for this legal phenomenon carries across jurisdiction, reasoning based on *prima facie* or circumstantial evidence is commonly associated with the doctrine of res *ipsa loquitur* ("the thing speaks for itself").[63] It is described this way because, in certain cases, the very nature of the event provides its own explanation on how that event should be evaluated. As the classics put it, "in the ordinary course of things, bags of flour do not fall from warehouse windows, stones are not found in buns, cars do not mount the pavement, and slippery substances are not left on shop floors."[64] Likewise, in the ordinary course of things, a bottle or boiler does not suddenly explode.[65] These situations are usually presented to reveal that sometimes the very occurrence of the event serves as strong *prima facie* evidence that something would not ordinarily happen without a defect.[66] In those cases, courts may, on that basis, presume defectiveness—and in some circumstances, also causation.

This reasoning is explicitly codified in the Directive. In particular, Article 10(2)(b) allows courts to presume defectiveness where a product fails to comply with mandatory product-safety requirements specifically intended to protect against the relevant risk, treating that non-compliance as *prima facie* sufficient. Likewise, Article 10(2)(c) applies when damage results from an "obvious malfunction" during "reasonably foreseeable" use, which the Directive considers sufficient to establish defectiveness. Accordingly, under this provision, if an autonomous cleaning robot suddenly accelerates and injures a user during normal operation, this obvious malfunction may allow the court to presume defectiveness without identifying the precise point of failure in the AI's control logic. The Directive extends this presumption-based logic to causation. Under Article 10(3), once defectiveness is

62 Vandenbussche (n 58) 20

63 Fairgrieve and Goldberg (n 14) 684.

64 Adrian Keane and Paul McKeown, *The Modern Law of Evidence* (9th edn, OUP 2012) 81.

65 W Page Keeton and others, *Prosser and Keeton on the Law of Torts* (5th edn, West Publishing Co 1984) 244-55.

66 ibid

established, causation may also be presumed where the type of damage is "typically consistent" with the defect in question, since certain harms are so characteristically linked to such defects that no further proof is required.[67] To illustrate, if an AI-operated drone has a known defect affecting navigational stability and subsequently crashes, the resulting property damage may be regarded as typically consistent with that defect, meaning the claimant need not provide additional evidence linking the defect to the crash.

7 CONCLUSIONS

The purpose of this article is doctrinal: to clarify the structure of the revised Directive by identifying its key features and novelties, and by analysing how its provisions work together as a coherent framework for resolving disputes involving product-caused damage, particularly harm resulting from AI products. To that end, the article organised the Directive's provisions into four categories. It began with the scope of application, defining when product liability applies. The second part examined the elements of liability, focusing on damage and defectiveness, and proposed a way to conceptualise the new "safety required by law" test alongside the traditional "consumer expectation" test in light of recent revisions. The third part addressed exemptions from liability, relevant to the defendant, explaining why such exemptions appear both in the dedicated article and elsewhere in the Directive. Finally, the article analysed the new structural addition to the product liability regime—the inclusion of procedural rules. Because of this fourth feature—procedural rules that ease the claimant's evidentiary burden—the Directive alters how EU product liability should be conceptually defined. Whereas the 1985 Directive rested on two conceptual axes that rendered the regime "strict"—replacing fault with product defectiveness and limiting defences—the revised Directive adds a third axis: burden-alleviation rules, which make liability even stricter in practice. Accordingly, if a proper definition of strict liability must consider all structural elements that contribute to its strictness, including this third axis.

REFERENCES

1. Alheit K, 'The Applicability of the EU Product Liability Directive to Software' (2001) 34 Comparative and International Law Journal of Southern Africa 188
2. Bruyne JD, Dheu O and Ducuing C, 'The European Commission's Approach to Extra-Contractual Liability and AI: An Evaluation of the AI Liability Directive and the Revised Product Liability Directive' (2023) 51 Computer Law & Security Review 105894. doi:10.1016/j.clsr.2023.105894
3. Calabresi G, *The Costs of Accidents: A Legal and Economic Analysis* (Yale UP 1970)

67 ibid

4. Cappelletti M, *Justifying Strict Liability: A Comparative Analysis in Legal Reasoning* (OUP 2022). doi:10.1093/oso/9780192859860.001.0001
5. Fairgrieve D and Goldberg R, *Product Liability* (3rd edn, OUP 2020). doi:10.1093/oso/9780199679232.001.0001
6. Fairgrieve D and others, 'Product Liability Directive' in Machnikowski P (ed), *European Product Liability: An Analysis of the State of the Art in the Era of New Technologies* (Intersentia 2016) 17
7. Fischer DA and others, *Products Liability: Cases and Materials* (West Academic Publishing 2022)
8. Gilead I, Koch BA and Green MD (eds), *Proportional Liability: Analytical and Comparative Perspectives* (De Gruyter 2013)
9. Howells G (ed), *Law of Product Liability* (2nd edn, LexisNexis Butterworths 2007)
10. Keane A and McKeown P, *The Modern Law of Evidence* (9th edn, OUP 2012)
11. Keeton WP and others, *Prosser and Keeton on the Law of Torts* (5th edn, West Publishing Co 1984)
12. Lohsse S and others, 'Liability for Artificial Intelligence' in Lohsse S, Schulze R and Staudenmayer D (eds), *Liability for Artificial Intelligence and the Internet of Things* (Nomos 2019) 9. doi:10.5771/9783845294797-9
13. Machnikowski P, 'Defectiveness' (The New Product Liability Directive: Doctrinal, Comparative and Interdisciplinary Approaches: Conference, Maastricht University, 17-18 October 2024)
14. Owen DG, *Products Liability in a Nutshell* (10th edn, West Academic Press 2023)
15. Owen DG, *Products Liability Law* (3rd edn, West Group 2015)
16. Pape SB, *Warnings and Product Liability: Lessons Learned from Cognitive Psychology and Ergonomics* (Eleven International Publishing 2012)
17. Parisi F and Frezza G, 'Burdens of Proof in Establishing Negligence: A Comparative Law and Economics Analysis' (2023) 9 Italian Law Journal 77
18. Rimkute D, 'AI Liability After the AILD Withdrawal: Why EU Law Still Matters?' (*Oxford Business Law Blog*, 1 April 2025) <https://blogs.law.ox.ac.uk/oblb/blog-post/2025/04/ai-liability-after-aild-withdrawal-why-eu-law-still-matters> accessed 7 June 2025
19. Smits JM, 'What is Legal Doctrine?: On the Aims and Methods of Legal-Dogmatic Research' in Gestel R, Micklitz HW and Rubin EL (eds), *Rethinking Legal Scholarship: A Transatlantic Dialogue* (CUP 2017) 207. doi:10.2139/ssrn.2644088
20. Taschner HC, 'Product Liability: Basic Problems in a Comparative Law Perspective' in Fairgrieve D (ed), *Product Liability in Comparative Perspective* (CUP 2005) 155. doi:10.1017/CBO9780511493850.011

21. Van Dam C, *European Tort Law* (3rd edn, OUP 2013)

22. Vandenbussche W, 'Dealing with Evidentiary Deficiency in Tort Law' (*SSRN*, 15 February 2019) doi:10.2139/ssrn.3335377

23. Verheyen T, 'Modern Theories of Product Warnings and European Product Liability Law' (2019) 15(3) Utrecht Law Review 44. doi:10.36633/ulr.541

24. Wagner G, 'Next Generation EU Product Liability – For Digital and Other Products' (2024) 15(2) Journal of European Tort Law 172. doi:10.1515/jetl-2024-0011

25. Wagner G, 'Software as a Product' in Lohsse S, Schulze R and Staudenmayer D (eds), *Smart Products* (Nomos 2022) 157. doi:10.5771/9783748929772-157

26. Wuyts D, 'The Product Liability Directive – More than Two Decades of Defective Products in Europe' (2014) 5(1) Journal of European Tort Law 1. doi:10.1515/jetl-2014-0001

AUTHORS INFORMATION

Deimantė Rimkutė
PhD student (Law), Junior Assistant, Faculty of Law, Vilnius University, Vilnius, Lithuania
deimante.rimkute@tf.vu.lt
https://orcid.org/0009-0002-2564-7487
Corresponding author, solely responsible for the manuscript preparing.

Competing interests: No competing interests were disclosed.

Disclaimer: The author declares that their opinion and views expressed in this manuscript are free of any impact of any organizations.

RIGHTS AND PERMISSIONS

EDITORS

ABOUT THIS ARTICLE

Cite this article
Rimkutė D, 'The New EU Product Liability Directive: Doctrinal Analysis' (2025) 8(Spec) Access to Justice in Eastern Europe 74-97 <https://doi.org/10.33327/AJEE-18-8.S-c000160>

DOI: https://doi.org/10.33327/AJEE-18-8.S-c000160

Keywords: *Product Liability Directive, AI liability, European tort law, product liability, AI and law, software liability.*

DETAILS FOR PUBLICATION

Date of submission: 31 Oct 2025
Date of acceptance: 05 Dec 2025
Online First Publication: 19 Dec 2025
Last Published: 30 Dec 2025
Whether the manuscript was fast tracked? - No
Number of reviewer report submitted in first round: 2 reports
Number of revision rounds: 1 round with conditional acceptance

Technical tools were used in the editorial process:
Plagiarism checks - Turnitin from iThenticate https://www.turnitin.com/products/ithenticate/
Scholastica for Peer Review https://scholasticahq.com/law-reviews

AI DISCLOSURE STATEMENT

The author confirms that AI technologies have only been used to enhance language clarity and grammar. No AI tools were used to generate ideas, structure arguments, analyze data, or produce conclusions.

АНОТАЦІЯ УКРАЇНСЬКОЮ МОВОЮ

Дослідницька стаття

НОВА ДИРЕКТИВА ЄС ПРО ВІДПОВІДАЛЬНІСТЬ ЗА ЯКІСТЬ ПРОДУКЦІЇ: ДОКТРИНАЛЬНИЙ АНАЛІЗ

Дейманте Рімкуте

АНОТАЦІЯ

Вступ. *Питання про те, хто несе відповідальність, коли система ШІ завдає шкоди, вже давно обговорюється в галузі ШІ та права. Коли ЄС запропонував Директиву про відповідальність за якість продукції та Директиву про відповідальність ШІ у 2022 році, багато хто очікував, що ці два документи спільно визначатимуть, як будуть розглядатися такі випадки. Ця ситуація змінилася після того, як Директива про відповідальність ШІ була відкликана в лютому 2025 року, а переглянута Директива про відповідальність за якість продукції була схвалена в жовтні 2024 року. У результаті ЄС тепер забезпечує гармонізовані правила відповідальності лише у випадках, коли продукти ШІ завдають шкоди споживачам або іншим фізичним особам. З огляду на цю зміну, актуальні дискусії щодо відповідальності ШІ можуть або розглянути наслідки відкликання Директиви про відповідальність ШІ, або уважно розглянути, як оновлена Директива (ЄС) 2024/2853 («Директива») розподіляє відповідальність. У цій статті використовується другий підхід. Що стосується споживачів та фізичних осіб, до яких застосовується Директива, стверджується, що дискусія має зосередитися на аналізі конкретної структури відповідальності, встановленої Директивою.*

Методи. *Розглянута Директива не є простим інструментом для тлумачення, зважаючи на її специфічну термінологію, високий рівень деталізації та розширені положення lex specialis порівняно із загальним принципом відповідальності за вчинене діяння. Хоча вона замінює Директиву 1985 року – і таким чином ґрунтується на наявному корпусі наукових досліджень – перегляд 2024 року вносить зміни, які вимагають повторного аналізу. Найкращим методом для розуміння розглянутого режиму відповідальності, уточнення його неоднозначностей, нововведень та проблемних аспектів є доктринальний правовий підхід. Відповідно, у цій статті використовується доктринальний аналіз для вивчення та систематизації зазначеної Директиви.*

Результати та висновки. *У статті положення Директиви поділено на чотири категорії: перша – сфера застосування, яка визначає, коли притягуються до відповідальності за якість продукції; друга – елементи відповідальності, які визначають, що позивач повинен довести; третя – захист, що вказує на винятки у сфері відповідальності, на які може посилатися відповідач; і, нарешті, процесуальні правила, які регулюють розкриття доказів, що стосуються обох сторін, та умови, за яких тягар доказування, що лежить на позивачі, може бути перекладений для вирішення доказових проблем.*

На основі цієї четвертої ознаки – процесуальних правил, що полегшують доказовий тягар позивача – у статті стверджується, що Директива ЄС змінює концептуальне визначення відповідальності за якість продукції. Згідно з Директивою 1985 року, ця структура ґрунтувалася на двох концептуальних засадах, що визначають сувору відповідальність за якість продукції: по-перше, заміна продукції з дефектом; та по-друге, обмеження захисту, обидва з яких зробили відповідальність за якість продукції «суворою». Оскільки нові процесуальні правила ще більше зміцнюють позицію позивача, у статті зроблено висновок, що у проаналізованій Директиві було додано третій пункт до концепції суворої відповідальності за якість продукції, що робить режим ще суворішим завдяки правилам полегшення тягаря доказування.

Ключові слова. *Директива про відповідальність за якість продукції, відповідальність штучного інтелекту, європейське деліктне право, відповідальність за якість продукції, штучний інтелект та право, відповідальність за програмне забезпечення.*

Access to Justice in Eastern Europe
ISSN 2663-0575 (Print)
ISSN 2663-0583 (Online)
http://ajee-journal.com

Research Article

ARTIFICIAL INTELLIGENCE IN COURTS AND DISPUTE RESOLUTION: CHALLENGES AND OPPORTUNITIES

Muhammad Qadeer Ashraf

ABSTRACT

Background: *Can AI be utilised in adjudication without compromising justice? While many support AI's potential to enhance judicial efficiency, concerns persist regarding its use in autonomous decision-making and the risks this poses to fundamental rights. This research assesses the transformative potential of artificial intelligence in the judicial sector and analyses the legal challenges associated with deploying AI in different capacities, including assistive roles, procedural functions, and the more controversial fully autonomous adjudication. It further explores existing legal frameworks and regulatory responses, with a particular focus on the EU AI Act, to assess how AI can be governed to balance efficiency with the protection of human rights and the rule of law.*

Methodology: *This research adopted a doctrinal research methodology, conducting a thorough analysis of primary sources, including the EU Artificial Intelligence Act, the European Convention on Human Rights (ECHR), and case law from US and Chinese jurisdictions. In addition, the study employed a comparative legal analysis to examine similarities and differences in AI regulation, judicial practices, and human*

DOI:
https://doi.org/10.33327/AJEE-18-8.S-a000152

Date of submission: 15 Jul 2025
Date of acceptance: 05 Oct 2025
Online First publication: 04 Dec 2025
Last Published: 30 Dec 2025

Disclaimer:

The author declares that his opinion and views expressed in this manuscript are free of any impact of any organizations.

rights protections across these jurisdictions. A systematic legal method was applied by analysing the EU AI Act in conjunction with ECHR provisions and relevant case law to evaluate the broader legal and regulatory framework governing AI in adjudication. Secondary sources, including journal articles, books, dissertations, conference papers, newspaper articles, reports, and blogs, were critically assessed using the CRAAP test. Finally, a critical legal method was applied to evaluate the ethical, procedural, and human rights implications of AI deployment in judicial decision-making.

Results and conclusion: *The findings of this research support the deployment of AI in an assistive capacity within adjudication, rather than as a fully autonomous decision-maker. AI-assisted adjudication can be strengthened through robust ethical and procedural safeguards, including mandatory human oversight, enhanced transparency, and improved interpretability of algorithmic decisions. Such measures can mitigate risks related to fairness, bias, and the erosion of judicial discretion identified in fully autonomous systems. When AI tools function explicitly as support rather than authoritative actors, concerns about violations of the right to a fair trial, judicial independence, and effective participation are significantly reduced. Ultimately, the legitimacy of AI in courtroom decision-making depends less on its mere presence and more on how it is designed, regulated, and perceived by both the judiciary and the public. Based on these findings, this research offers concrete recommendations for judicial institutions to ensure that AI enhances efficiency without compromising justice.*

1 INTRODUCTION

AI is transforming justice worldwide. In 2019, a Chinese litigant attended a trial entirely online—with evidence, witnesses, a judge, and a ruling all conducted through a virtual platform guided by AI. In that proceeding, the human judge relied not only on digital records but also on AI-generated case summaries and precedent recommendations.[1] The verdict was delivered within minutes.

Half a world away, a defendant in Wisconsin, USA, filed a motion for post-conviction relief after learning that the original decision had been influenced by a proprietary algorithm (COMPAS) that neither they nor their counsel could question or understand.[2] When challenged, the software's creators refused to disclose how the system made its decisions, claiming trade secrecy.

Similarly, in Canada, the federal Post-Conviction Risk Assessment tool—used to inform probation decisions—was shown to produce racially disparate results, assigning

1 Xinhua, 'Beijing Internet Court Launches AI Judge' *China Daily* (Beijing, 28 June 2019) <https://www.chinadaily.com.cn/a/201906/28/WS5d156cada3103dbf1432ac74.html> accessed 23 May 2025.

2 *State v Loomis* 881 NW2d 749 (Wis 2016).

disproportionately higher risk scores to Black offenders compared to white offenders.[3] These cases highlighted the changing role of AI in the justice system: one driven by digital convenience, the other shadowed by intentional and inherent opacities.[4]

In Europe, Estonia briefly became the subject of international media speculation in 2019 when reports claimed it was developing a "robot judge" to decide small claims disputes. The Estonian government later clarified that no such fully autonomous judicial project was underway and that its efforts were limited to exploring digital tools to improve efficiency in the justice system.[5] This episode illustrates how quickly narratives concerning AI in adjudication can drift from reality, feeding public anxieties about automation in the judiciary.

Yet all four cases reflect a deeper transformation quietly underway in courtrooms across the world. AI is no longer a speculative concept in law; it has become a practical tool used to evaluate evidence, predict outcomes, streamline caseloads, and increasingly, to assist in adjudication. From automated case sorting in the Netherlands to predictive analytics in UK asylum rulings, and from Tax Foresight in Canada to algorithmic mediation in Singapore's small claims tribunals, AI is beginning to shape who receives justice, when, and on what terms.

But with this transformation comes a set of paradoxes. While AI is expected to resolve judicial backlogs and democratise legal access, it also presents significant challenges to fundamental rights: human dignity, fairness, and open justice.

2 METHODOLOGY

This research adopted a doctrinal legal methodology to examine the integration of artificial intelligence in European judicial systems, specifically in courts and dispute resolution processes. It addresses growing scholarly and regulatory concerns about whether AI can enhance the delivery of justice without compromising foundational legal values, including judicial independence, transparency, human dignity, and the right to a fair trial. The study is particularly timely, given the recent proliferation of AI tools designed to assist or, in some cases, replace judicial decision-making in Europe. The following core objectives guide the research:

3 Jennifer L Skeem and Christopher T Lowenkamp, 'Risk, Race, and Recidivism: Predictive Bias and Disparate Impact' (2016) 54(4) Criminology 685-700. doi:10.1111/1745-9125.12123.

4 Jenna Burrell, 'How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms' (2016) 3(1) Big Data & Society 1-2. doi:10.1177/2053951715622512.

5 Maria-Elisa Tuulik, 'Estonia Does Not Develop AI Judge' (*Ministry of Justice and Digital Affairs, Republic of Estonia*, 16 February 2022) <https://www.justdigi.ee/en/news/estonia-does-not-develop-ai-judge> accessed 23 May 2025.

- To examine how far the use of AI in court decision-making aligns with legal principles supporting justice in Europe.
- To assess the extent to which artificial intelligence can replace human judges in judicial systems.
- To evaluate whether the integration of AI into judicial processes can enhance the quality and effectiveness of justice delivery without undermining core legal values.

To achieve these objectives, the study conducted a thorough analysis of primary sources, including the EU Artificial Intelligence Act, the European Convention on Human Rights (ECHR), and case law from US and Chinese jurisdictions, with particular attention to Articles 6, 13, 14, and 45 of the ECHR. It also examined relevant domestic and international soft law instruments, such as the CEPEJ Ethical Charter, the OECD AI Principles, and UNESCO's Recommendation on the Ethics of AI, selected for their influence on the EU AI Act and their guidance on court-level practices.

Secondary sources—including journal articles, books, dissertations, conference papers, newspaper articles, reports, and blogs—were critically analysed to provide context, assess current scholarly debates, and identify normative gaps. These sources were collected from legal databases and academic repositories and evaluated using the CRAAP test.

The methodology incorporated multiple complementary approaches: A comparative legal analysis to examine regulatory and judicial frameworks across Europe, the US, and China, highlighting similarities, divergences, and lessons learned. A systematic legal method, analysing the EU AI Act alongside ECHR provisions and relevant case law to provide a structured assessment of legal norms, principles, and obligations. A critical legal method, assessing the ethical, procedural, and human rights implications of AI in judicial decision-making, identifying risks, and proposing normative and regulatory recommendations. This layered approach, combining doctrinal, comparative, systematic, and critical methods, supported an interpretive, evaluative, and normative analysis: interpreting legal norms and jurisprudence, comparing institutional frameworks across jurisdictions, and evaluating the adequacy of current legal regimes.

While no empirical data collection was conducted, illustrative global case references were used to contrast regulatory approaches and highlight practical concerns. This methodology enables the research to provide a well-reasoned, independent legal perspective on the role of AI in judicial systems.

3 RESULTS AND DISCUSSION

AI systems are increasingly permeating the judicial sector worldwide—not only in terms of digital case management but also in supporting core legal tasks such as legal reasoning, data retrieval, and procedural structuring. While the use of AI as a substitute for judicial

decision-making—popularly referred to as "robot judges"—remains highly contested, its application in a supportive or assistive role has gained greater legitimacy in contemporary legal research. As recognised and encouraged by the CEPEJ, such supportive uses of AI can contribute to procedural efficiency, broaden access to justice, and strengthen institutional accountability.

3.1. AI's Transformative Potential in the Judicial Sector

As AI systems evolve, their potential to streamline courtroom procedures becomes more apparent. In its assistive role, AI can contribute to a wide range of courtroom and pre-trial tasks,[6] including drafting judgments, intelligent case routing, transcription and translation of legal documents, anonymisation of judgments, and integration with national e-Government systems for the verification of documentary evidence.

Particularly notable is AI's potential in managing repetitive, low-value claims through online dispute resolution (ODR) platforms[7]—a development that can reduce case backlog without excluding legal recourse.[8] Moreover, AI can assist judges in technical domains, such as the calculation of financial penalties, the evaluation and distribution of property in family disputes, and the identification of applicable statutes or precedents through efficient legal research algorithms. For example, natural language processing (NLP) models can be used to analyse the factual matrix of cases, and those can suggest relevant legal provisions with remarkable speed and precision.[9]

While these tools do not exercise discretion, they enhance the accuracy and consistency of judicial outputs, contributing to the predictability of legal decisions—an important component of legal certainty and the rule of law. The breadth of AI applications in LegalTech is supported by recent data from the CEPEJ-AIAB, which reported that 125 AI and cyberjustice tools[10] are currently in use or undergoing testing across various European jurisdictions.

6 Ignacio N Cofone, 'AI and Judicial Decision-Making' in Florian Martin-Bariteau and Teresa Scassa (eds), *Artificial Intelligence and the Law in Canada* (LexisNexis Canada 2021) ch 13, 8.

7 Tania Sourdin, 'Judge v Robot? Artificial Intelligence and Judicial Decision-Making' (2018) 41(4) UNSW Law Journal 1114; Tania Sourdin and others, 'COVID-19, Technology and Family Dispute Resolution' (2020) 30(4) Australasian Dispute Resolution Journal 270.

8 Jessica Rosberger, 'AI Mediation for Reducing Court Congestion' (26 November 2024) Cornell Journal of Law and Public Policy <https://publications.lawschool.cornell.edu/jlpp/2024/11/26/ai-mediation-for-reducing-court-congestion/> accessed 22 May 2025.

9 Harry Surden, 'Artificial Intelligence and Law: An Overview of Recent Technological Changes in Large Language Models and Law' (2025) 96 University of Colorado Law Review 376. doi:10.2139/ssrn.5135305.

10 CEPEJ, '1st AIAB Report on the Use of Artificial Intelligence (AI) in the Judiciary Based on the Information Contained in the Resource Centre on Cyberjustice and AI' (CEPEJ-AIAB(2024)4Rev5, 28 February 2025) <https://rm.coe.int/cepej-aiab-2024-4rev5-en-first-aiab-report-2788-0938-9324-v-1/1680b49def> accessed 2 June 2025.

These tools, in turn, present distinct benefits and regulatory challenges. For instance, tools supporting document discovery and legal research undoubtedly enhance efficiency but can also amplify existing biases if not regularly audited. Similarly, ODR platforms improve accessibility but raise fairness concerns when human oversight is minimal. Moreover, AI-based predictive tools can support in organising and prioritising cases; however, their heavy reliance on previous datasets risks uniform adjudication, which can undermine individualised justice.[11] In addition, decision-support systems that propose sentencing or summarise case facts should augment, not substitute, judicial reasoning; otherwise, they risk diminishing the role of human conscience in adjudication. At the same time, anonymisation tools help safeguard data privacy but can inadvertently hinder transparency in legal scholarship and appellate review. Equally important, translation and transcription tools must be highly accurate, as even slight misinterpretations can lead to unfair outcomes.

Nowadays, NLP models are also reshaping the broader landscape of legal practice. Various law firms already employ either customised or general generative artificial intelligence (GenAI) tools.[12] Between 2023 and 2024, these models became increasingly recognised for their ability to generate human-like text, making them valuable assets for tasks such as legal document drafting, contract analysis, and the provision of preliminary legal advice.[13] Additionally, the integration of GenAI has expanded the range of AI-supported functions in legal workflows, including summarising complex legal content, producing tailored documents, and facilitating interaction through question-answer systems.

Open-source generative AI platforms, such as those evaluated by French institutions,[14] are also contributing to increased transparency and usability by offering clearer insights into the architecture and governance of generative models. In the private legal sector, GenAI tools are already being explored for their potential to improve productivity and reduce routine workload. Although deployment within judicial institutions remains cautious due to higher ethical and procedural standards, the commercial legal sector has seen a growing interest in adopting these systems for case preparation, legal research, and client communication. As user involvement in system design and testing increases, generative AI is likely to become a practical instrument for enhancing the speed, efficiency, and accessibility of justice delivery.

Although these tools are efficient and automate court procedures, this automation does not come at the cost of due process.

11 Bhishm Khanna, *Predictive Justice: Using AI for Justice* (Centre for Public Policy Research 2021) 6.

12 Jonathan Kewley and others, 'Fast Law: Why Speed is the Priority for Lawyers Using AI' (*LexisNexis*, 2024) <www.lexisnexis.co.uk/insights/fast-law-why-speed-is-the-priority-for-lawyers-using-ai/index.html> accessed 2 June 2025.

13 David Uriel Socol de la Osa and Nydia Remolina, 'Artificial Intelligence at the Bench: Legal and Ethical Challenges of Informing—or Misinforming—Judicial Decision-Making through Generative AI' (2024) 6 Data & Policy e59. doi:10.1017/dap.2024.53.

14 PEReN, 'Open Source GenAI Comparator' (*Government of France, PEReN - Centre of Expertise for Digital Platform Regulation*, 2025) <https://www.peren.gouv.fr/en/compare-os-iag/> accessed 29 April 2025.

3.2. Ethical Guidelines for AI in the Justice System

Moving forward, it is important to consider the ethical standards guiding AI's integration into the judicial process. The European Ethical Charter on the use of AI in judicial systems, adopted in 2018, articulates five ethical principles to guide the responsible integration of AI, often referred to as LegalTech[15] within judicial institutions.[16] These principles, though not binding, have emerged as soft law instruments that have influenced both national legislation and court-level policy reforms across Europe.

As the integration of AI tools in the judicial process becomes more widespread, attention must be paid to their ethical governance. Although AI tools promise significant procedural benefits, their deployment must remain aligned with constitutional principles. As judicial reliance on algorithmic systems increases, legal systems must simultaneously reinforce mechanisms for auditability, human review, and accountability.

While current AI systems cannot substitute for judicial conscience or legal reasoning, they can serve as valuable co-pilots by streamlining processes, improving institutional performance, and broadening access to legal remedies when governed ethically.[17]

AI technologies are being deployed in adjudication in various ways: as *assistive tools* supporting judges, *procedural aids* managing administrative or evidentiary processes, and, more controversially, in *fully autonomous* roles that aim to replace human judicial discretion altogether. While each of these applications raises legal and ethical questions, the risks intensify as AI moves from assistive to fully autonomous functions.[18]

3.3. Assistive AI in Judicial Decision-Making

Despite their potential, assistive AI systems show varied levels of prediction accuracy. While deep learning and large language model–based tools can achieve around 85–90% accuracy on benchmark datasets for tasks such as charge prediction and legal article recommendation, their performance drops to 70–80% in more complex tasks like sentencing and multi-label judgment prediction.[19]

15 Z Seldağ Güneş Peschke and Lutz Peschke, 'Artificial Intelligence and the New Challenges for EU Legislation' (2022) 2 Yıldırım Beyazıt Hukuk Dergisi 1278. doi:10.33432/ybuhukuk.1104344.

16 CEPEJ, *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment* (Council of Europe 2018) <https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c> accessed 29 April 2025.

17 Filippo Donati, 'The Use of Artificial Intelligence in Judicial Systems: Ethics and Efficiency' in Mireia Artigot i Golobardes and others (eds), *Artificial Intelligence, Judicial Decision-Making and Fundamental Rights* (2nd edn, Scuola Superiore della Magistratura 2024) 15.

18 UNESCO, *Draft Guidelines for the Use of AI Systems in Courts and Tribunals* (CI/DIT/2025/GL/01, May 2025) <https://unesdoc.unesco.org/ark:/48223/pf0000393682> accessed 25 September 2025.

19 Chuyue Zhang and Yuchen Meng, 'Bridging the Divide: Technical Research and Application on Legal Judgment Prediction' [2025] Artificial Intelligence and Law. doi:10.1007/s10506-025-09473-7.

At the same time, this integration of AI in court processes raises substantial human rights risks, particularly when individuals have no viable alternatives.[20] Although the idea of AI assisting judges in sentencing, bail, or probation decisions offers greater efficiency in justice delivery, it can pose significant threats to the right to a fair trial,[21] due process, the right to an effective remedy, transparency, and protection against discrimination, while undermining the obligation of courts to provide reasoned judgments.[22]

AI systems used as assistive risk assessment tools are only as fair as the data on which they are trained. When fed historical data embedded with structural biases, these systems are prone to replicating and even amplifying those injustices. The phenomenon of the "feedback loop" further exacerbates the discriminatory potential of AI-driven judicial systems. When sentencing algorithms are trained on biased historical data, they can disproportionately impose harsher penalties on certain demographic groups. As these biased precedents are continuously fed back into the system for retraining,[23] they reinforce and perpetuate existing discriminatory patterns.[24]

Unlike algorithms, human judges can identify and rectify errors, prevent the same mistakes from being repeated in future adjudications, and be held accountable for discriminatory practices.[25] This danger is well illustrated in the case of *State of Kansas v. John Keith Walls* (2017), where the appellate court held that the defendant must be granted full access to the LSI-R (Level of Service Inventory-Revised) assessment, which the lower court had relied upon to determine the conditions of his probation. Denying the defendant access to this assessment prevented him from disputing the accuracy of information that played a critical role in the adjudication,[26] thereby violating his constitutional right to procedural due process.

By citing the case *Kansas v. Easterling*, the court concluded that the district court's refusal to provide the complete LSI-R report violated the defendant's constitutional right to

20 Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the Human Rights Impacts of Algorithmic Systems (8 April 2020) Preamble, para 11 <Https://Search.Coe.Int/Cm?I=09000016809e1154> Accessed 29 April 2025.

21 Charter of Fundamental Rights of the European Union [2012] OJ C 326/391, art 47.

22 Convention for the Protection of Human Rights and Fundamental Freedoms (4 November 1950) [1955] UNTS 213/222, arts 6, 13, 14 and 45.

23 Justice GC Martin, 'How Far has Technology Invaded the Criminal Justice System?' (ANZELA Legal Studies Teachers' Conference, Brisbane, 11 May 2018) 20 <https://www.sclqld.org.au/catalogue/records/89130> accessed 5 May 2025.

24 European Union Agency for Fundamental Rights, *#BigData: Discrimination in data-Supported Decision Making* (FRA 2018) <https://fra.europa.eu/en/publication/2018/bigdata-discrimination-data-supported-decision-making> accessed 5 May 2025.

25 Frank Pasquale and Glyn Cashwell, 'Prediction, Persuasion, and the Jurisprudence of Behaviourism' (2018) 68(1) University of Toronto Law Journal 66. doi:10.3138/utlj.2017-0056. See also: Lord Sales, 'Algorithms, Artificial Intelligence and the Law' (2020) 25(1) Judicial Review 46. doi:10.1080/10854681.2020.1732737.

26 *State of Kansas v John Keith Walls* no 116027 [2017] Kan Ct App.

procedural due process during the sentencing stage of the criminal proceedings.[27] Such risks show how even assistive AI can compromise fairness when defendants cannot understand or contest algorithmic reasoning.

3.4. Procedural AI in Court Administration

AI is also increasingly used in procedural roles, such as managing evidence, facilitating e-discovery, scheduling hearings, and providing translation or interpretation. While these applications appear less controversial, they directly affect the right to effective participation in proceedings. For instance, effective participation, as emphasised consistently by the ECtHR, involves more than physical presence; it requires that the accused fully comprehend, respond to, and contest the proceedings. If an AI translation system fails to provide accurate interpretation, or if automated scheduling denies adequate preparation time, the fairness of the trial is compromised. Even seemingly minor procedural issues, such as inadequate courtroom acoustics,[28] can impede participation. Procedural AI must therefore be critically evaluated to ensure that its efficiency gains do not come at the expense of defendants' rights.

3.5. Fully Autonomous AI ("Robot Judges")

The most serious challenges arise when AI is deployed as a fully autonomous adjudicator, replacing human judicial discretion. While human judges are entrusted with evaluating factors such as an offender's intent, remorse, and socio-personal circumstances,[29] AI systems are inherently devoid of emotional intelligence and contextual sensitivity.[30] By relying exclusively on quantitative metrics and pre-programmed logic,[31] such systems risk reducing the rich complexity of human conduct into inflexible algorithmic determinations.[32] For instance, an AI judge could impose an identical sentence on a coerced, first-time juvenile offender and a habitual adult recidivist, merely because the statutory elements of their offences align.[33] This mechanistic approach, lacking the capacity to discern the moral gravity or mitigating factors of each case, not only generates

27 *State of Kansas v David E Easterling* no 100,454 [2009] 213 P.3d 418 Kan Sup Ct.

28 Aleš Završnik, 'Criminal Justice, Artificial Intelligence Systems, and Human Rights' (2020) 20 ERA Forum 576. doi:10.1007/s12027-020-00602-0.

29 Richard Susskind, *Online Courts and the Future of Justice* (OUP2019) 206-7.

30 Benjamin Alarie, Anthony Niblett and Albert H Yoon, 'How Artificial Intelligence will Affect the Practice of Law' (2018) 68(1) University of Toronto Law Journal 108-9. doi:10.3138/utlj.2017-0052.

31 Australian Government, *Automated Assistance in Administrative Decision Making: Better Practice Guide* (Office of the Privacy Commissioner 2007) 4.

32 Robert J Condlin, 'Online Dispute Resolution: Stinky, Repugnant, or Drab' (2017) 18(3) Cardozo Journal of Conflict Resolution 723.

33 Isaac Taylor, 'Justice by Algorithm: The Limits of AI in Criminal Sentencing' (2023) 42(3) Criminal Justice Ethics 193. doi:10.1080/0731129X.2023.2275967.

disproportionate and potentially unjust outcomes but also undermines judicial discretion, proportionality, and substantive equality before the law.

The principle of judicial independence is also directly threatened by fully autonomous adjudication. Robot judges can inadvertently shift discretion from judges to engineers and programmers who design these systems, thereby hollowing out the independence of the judiciary. In such scenarios, the tribunal is no longer "independent" in substance, even if formally constituted under law. The independence and neutrality of verdicts rendered by robot judges can be jeopardised by even minor changes in the datasets on which they rely; a slight modification by a programmer could entirely alter the outcome of a case. Such vulnerability undermines the very concept of an impartial and independent tribunal.[34]

A particularly pressing concern is the absence of explainability in algorithmic decision-making. AI tools often function as opaque "black boxes",[35] generating outcomes that are neither transparent nor comprehensible to defendants or their legal counsel.[36] This deficiency contravenes the core principles of adversarial proceedings and the equality of arms by placing the defence at a marked disadvantage.[37] Judgments rendered by robot judges risk violating Article 45 of the ECHR, which mandates that "reasons shall be given for judgments as well as for decisions declaring applications admissible or inadmissible." If a convicted person cannot comprehend the *ratio decidendi*, the legal reasoning underpinning the judgment, they are effectively deprived of the opportunity to exercise their right of appeal.[38] This directly threatens Article 2 of Protocol No. 7 of the ECHR, which guarantees that "everyone convicted of a criminal offence by a tribunal shall have the right to have his conviction or sentence reviewed by a higher tribunal." Without a clear understanding of how and why a decision was reached, the exercise of this right becomes illusory, eroding one of the fundamental safeguards of the criminal justice system.

Courts occupy a uniquely sensitive role in upholding the rule of law and protecting fundamental rights. Unlike AI applications in commercial or administrative domains, judicial AI must operate in strict conformity with the constitutional guarantees enshrined

34 *Hermi v Italy* App no 18114/02 (ECtHR, 18 October 2006).

35 Ashley Deeks, 'The Judicial Demand for Explainable Artificial Intelligence' (2019) 119(7) Columbia Law Review 1833.

36 Monika Zalnieriute and Felicity Bell, 'Technology and Judicial Role' in Gabrielle Appleby and Andrew Lynch (eds), *The Judge, the Judiciary and the Court: Individual, Collegial and Institutional Judicial Dynamics in Australia* (CUP 2021) 116. doi:10.1017/9781108859332.

37 Council of Europe MSI-NET, 'Study on the Human Rights Dimensions of Automated Data Processing Techniques (in Particular Algorithms) and Possible Regulatory Implications' (MSI-NET(2016)06 rev3 FINAL, 6 October 2017) 10-1 <https://rm.coe.int/study-hr-dimension-of-automated-data-processing-incl-algorithms/168075b94a> accessed 30 April 2025.

38 Harry Surden, 'The Ethics of AI in Law: Basic Questions' in Markus D Dubber, Frank Pasquale and Sunit Das (eds), *The Oxford Handbook of Ethics of AI* (online edn, Oxford Academic 2020) 730-2. doi:10.1093/oxfordhb/9780190067397.013.46.

in regional legal frameworks.[39] The deployment of AI in this context, where decisions can directly impact fundamental rights and judicial independence, demands legal instruments that extend beyond general-purpose technological governance.[40] Notably, empirical data from the AI on Trial project indicates that 46 disputes involving the use of AI technologies across various sectors have already reached European courts.[41] In some instances, overly restrictive or innovation-averse judgments have risked stifling the development of beneficial AI tools,[42] including those designed for legal research, document management, and procedural efficiency. These developments highlight the urgent need for a coherent, robust legal framework to govern the use of AI across all sectors, and particularly within the judiciary.

3.6. EU AI Act (2024) and Soft Laws

The recent EU AI Act, the first comprehensive regulation on AI, adopts a structured risk-based regulatory model, under which AI systems intended for use in the administration of justice are designated as "high-risk".[43] This classification is not arbitrary; it reflects the recognition within the Act that such systems can produce flawed predictions,[44] introduce biases, and even generate misleading[45] or hallucinated content[46] when applied to legal reasoning or case-specific recommendations.

39 Paweł Marcin Nowotko, 'AI in Judicial Application of Law and the Right to a Court' (2021) 192 Procedia Computer Science 2220. doi:10.1016/j.procs.2021.08.235.

40 Giulia Gentile, 'Trial by Artificial Intelligence: How Technology Is Reshaping Our Legal System' (*LSE - London School of Economics and Political Science*, 8 September 2023) <https://blogs.lse.ac.uk/politicsandpolicy/trial-by-artificial-intelligence-how-technology-is-reshaping-our-legal-system/> accessed 13 May 2025.

41 Isadora Valadares Assunção, 'Beyond Regulation: What 500 Cases Reveal about the Future of AI in the Courts' (*TechPolicy Press*, 20 May 2025) <https://www.techpolicy.press/beyond-regulation-what-500-cases-reveal-about-the-future-of-ai-in-the-courts/> accessed 21 May 2025.

42 Neel Guha, Peter Henderson and Diego A Zambrano, 'Gamesmanship in Modern Discovery Tech' in David Freeman Engstrom (ed), *Legal Tech and the Future of Civil Justice* (CUP 2023) 112. doi:10.1017/9781009255301.008.

43 Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), para 26 <http://data.europa.eu/eli/reg/2024/1689/oj> accessed 30 April 2025.

44 ibid, para 61.

45 *Mark Walters v OpenAI* no 23-A-04860-2 [2025] Ga Super Ct.

46 *Mata v Avianca Inc* no 22-CV-1461 [2023] 678 F Supp. 3d 443 SDNY. See also: James E Dority, Christian E Mammen and Jill Rothstein, 'A "Brief" Hallucination by Generative AI Can Land You in Hot Water' (*Mondaq*, 15 June 2023) <https://www.mondaq.com/unitedstates/new-technology/1329860/a-brief-hallucination-by-generative-ai-can-land-you-in-hot-water> accessed 22 May 2025.

These concerns are not merely technical; they implicate foundational legal principles such as legal certainty, procedural fairness, and the right to a fair trial (as mentioned above). By explicitly acknowledging the serious impact judicial AI can have on democratic institutions and individual rights, the regulation signals that the administration of justice requires heightened vigilance where automation is introduced.[47]

In doing so, the Act draws an important conceptual line between AI tools that influence judicial decision-making and those that serve administrative or ancillary purposes. AI systems that assist in the interpretation of facts, the application of law, or the adjudication of disputes—whether in formal courts or ADR mechanisms—are subjected to greater regulatory oversight, particularly when their outputs bear legal consequences. This distinction reflects a principled commitment to preserving human discretion within the judicial process.[48]

The regulation appears to endorse the use of AI as a supportive instrument, but not as a surrogate for judicial conscience or interpretive reasoning.[49] Notably, the Act exempts systems designed solely for administrative tasks, such as anonymisation or internal document handling, from the high-risk category, demonstrating regulatory proportionality and avoiding undue constraint on harmless innovation.[50]

Moreover, the AI Act explicitly embodies the normative ethos of earlier soft law instruments such as the EDDRP,[51] the CEPEJ-AIAB guidelines, and the recommendations of the EU AI-HLEG,[52] despite their non-binding nature. Although these bodies operate independently, they converge around key ethical principles applicable to high-risk AI systems, particularly those deployed in the justice sector.

The principle of human agency and oversight, consistently emphasised in soft law frameworks, is mirrored in the AI Act's requirement that final legal determinations remain within the remit of human judges. By mandating human oversight, the AI Act acknowledges that judicial independence is not merely formal; it is functional and interpretive, demanding

47 Christoph K Winter, 'The Challenges of Artificial Judicial Decision-Making for Liberal Democracy' in Piotr Bystranowski, Bartosz Janik and Maciej Próchnicki (eds), *Judicial Decision-Making: Economic Analysis of Law in European Legal Scholarship* (Springer Cham 2022) 179. doi:10.1007/978-3-031-11744-2_9.

48 Socol de la Osa and Remolina (n 13) e59-21.

49 Benjamin Minhao Chen, Alexander Stremitzer and Kevin Tobia, 'Having Your Day in Robot Court' (2022) 36(1) Harvard Journal of Law & Technology 168.

50 Artificial Intelligence Act (n 43) para 61.

51 European Declaration on Digital Rights and Principles for the Digital Decade (15 December 2022) <https://digital-strategy.ec.europa.eu/en/library/european-declaration-digital-rights-and-principles> accessed 30 April 2025.

52 High-Level Expert Group on Artificial Intelligence, *Ethics Guidelines for Trustworthy AI* (European Commission 2019) 26-31 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> accessed 30 April 2025.

sustained cognitive engagement from the judge.[53] Overreliance on AI-generated recommendations risks diminishing the judge's active role in legal interpretation and could, over time, degrade the very faculties upon which just adjudication depends.[54]

Thus, the hybrid adjudication model endorsed by the AI Act—a collaborative interface between human judgment and algorithmic assistance—aims not only to safeguard legality but also to preserve public confidence in the judiciary.[55] A first-ever international legally binding treaty on AI under the Council of Europe has already been signed by 46 countries.[56]

3.7. International Legal Instruments

Global soft law instruments play a crucial role in the governance of AI, complementing hard-law frameworks. Though non-binding, these principles have significantly informed legislative agendas and regulatory discourse, particularly within jurisdictions committed to fundamental rights, democratic governance, and responsible innovation.

A widely accepted framework outlines five key principles that underpin the concept of trustworthy AI.[57] From a judicial perspective, these principles collectively form a normative foundation that complements European regulatory instruments. The emphasis on transparency and explainability addresses the legal need to understand how AI systems reach decisions. By promoting human oversight, the framework safeguards judicial independence amid increasing automation. The principle of robustness and accountability principles further advocate traceability and risk assessments to uphold fairness. Together, the OECD principles serve as a coherent ethical blueprint, reinforcing the broader international consensus around the responsible use of AI.

Another international framework presents ten core principles that demonstrate an effort to embed human rights, sustainability, and democratic accountability into the full lifecycle of AI systems.[58] Even though these recommendations are non-binding, they establish universally accepted criteria for the use of AI in an ethical manner, especially in the judicial sector, where the use of technology must respect fundamental human rights. Of particular importance is the principle of proportionality and the mandate to "do no harm," which

53 Michael E Donohue, 'A Replacement for Justitia's Scales? Machine Learning's Role in Sentencing' (2019) 32(2) Harvard Journal of Law & Technology 672.

54 Cofone (n 6) 7.

55 Chen, Stremitzer and Tobia (n 49) 163.

56 'US, Britain, EU to Sign First International AI Treaty' (*Reuters,* 6 September 2024) <https://www.reuters.com/technology/artificial-intelligence/us-britain-eu-sign-agreement-ai-standards-ft-reports-2024-09-05/> accessed 21 May 2025.

57 OECD, *Recommendation of the Council on Artificial Intelligence* (OECD/LEGAL/0449, OECD Legal Instruments 2025) 8-9 <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449> accessed 21 May 2025.

58 UNESCO, *Recommendation on the Ethics of Artificial Intelligence* (SHS/BIO/PI/2021/1, UNESCO 2022) 20-3 <https://unesdoc.unesco.org/ark:/48223/pf0000381137> accessed 21 May 2025.

echoes the European risk-based regulatory approach by requiring that AI uses remain strictly necessary to achieve legitimate objectives.

Similarly, the emphasis on human oversight and determination reinforces the imperative that automated systems must never substitute for human judicial discretion but rather remain subordinate to it. In judicial contexts, this ensures that the legitimacy of decisions remains rooted in legal reasoning rather than algorithmic inference. Transparency, explainability, and accountability also form core elements of the UNESCO framework. These principles are especially salient for courts, as they ensure that litigants can understand, interrogate, and, where necessary, challenge AI-influenced outcomes, an essential safeguard for preserving adversarial proceedings and the right to an effective remedy and trial.

Moreover, the Recommendation's insistence on fairness and non-discrimination reflects the growing awareness that AI systems can entrench or amplify structural biases if left unchecked, particularly in areas such as criminal sentencing, bail, and asylum determinations. The inclusion of sustainability as an evaluative criterion further expands the normative scope of AI ethics, underscoring the responsibility of AI actors to consider not only immediate harms but long-term legal and societal consequences.

4 CONCLUSIONS

Undoubtedly, AI has emerged as a powerful tool for enhancing judicial efficiency, and many jurisdictions have begun integrating AI into courts, primarily in supportive and assistive roles. While these applications show significant potential, they also present considerable legal and procedural challenges. Even when AI is used in an assistive capacity—particularly in tasks such as natural language processing (NLP) and generative AI—issues remain, including the production of inaccurate or legally misleading information. Although human judges can correct such errors, their recurring nature highlights the current fragility and limitations of AI systems.

The risks posed by AI differ depending on its role. In an assistive capacity, AI can support judicial decision-making, provided human oversight, transparency, and interpretability are maintained. In contrast, the deployment of AI in a fully autonomous capacity raises substantial legal concerns. Autonomous AI decision-making could compromise core principles of European law, including the right to a fair trial, judicial independence, the delivery of reasoned judgments, the right to appeal, and principles such as transparency, non-discrimination, and equality before the law.

These risks render the full replacement of human judges by AI both problematic and potentially inconsistent with the rule of law—although future improvements in AI design, explainability, and regulatory oversight may mitigate some of these concerns. Current frameworks, including the EU AI Act, still lack clear guidance for judicial actors on how to

integrate AI outputs into decision-making responsibly. This ambiguity creates gaps in algorithmic accountability, explainability, and procedural fairness.

To address these gaps, judicial councils and legal institutions should establish comprehensive guidelines defining when, and to what extent, AI-generated information may be relied upon in courts. Until AI systems can consistently meet the standards of legal reasoning, accountability, and human dignity, their role in adjudication should remain strictly assistive. Fully autonomous AI decision-making, if pursued at all, should be approached cautiously, with risks carefully managed.

REFERENCES

1. Alarie B, Niblett A and Yoon AH, 'How Artificial Intelligence will Affect the Practice of Law' (2018) 68(1) University of Toronto Law Journal 106. doi:10.3138/utlj.2017-0052
2. Assunção IV, 'Beyond Regulation: What 500 Cases Reveal about the Future of AI in the Courts' (*TechPolicy Press*, 20 May 2025) <https://www.techpolicy.press/beyond-regulation-what-500-cases-reveal-about-the-future-of-ai-in-the-courts/> accessed 21 May 2025
3. Burrell J, 'How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms' (2016) 3(1) Big Data & Society 1. doi:10.1177/2053951715622512
4. Chen BM, Stremitzer A and Tobia K, 'Having Your Day in Robot Court' (2022) 36(1) Harvard Journal of Law & Technology 127
5. Cofone IN, 'AI and Judicial Decision-Making' in Martin-Bariteau F and Scassa T (eds), *Artificial Intelligence and the Law in Canada* (LexisNexis Canada 2021) ch 13
6. Condlin RJ, 'Online Dispute Resolution: Stinky, Repugnant, or Drab' (2017) 18(3) Cardozo Journal of Conflict Resolution 717
7. Deeks A, 'The Judicial Demand for Explainable Artificial Intelligence' (2019) 119(7) Columbia Law Review 1829
8. Donati F, 'The Use of Artificial Intelligence in Judicial Systems: Ethics and Efficiency' in Artigot i Golobardes M and others (eds), *Artificial Intelligence, Judicial Decision-Making and Fundamental Rights* (2nd edn, Scuola Superiore della Magistratura 2024) 15
9. Donohue ME, 'A Replacement for Justitia's Scales? Machine Learning's Role in Sentencing' (2019) 32(2) Harvard Journal of Law & Technology 657
10. Dority JE, Mammen CE and Rothstein J, 'A "Brief" Hallucination by Generative AI Can Land You in Hot Water' (*Mondaq*, 15 June 2023) <https://www.mondaq.com/unitedstates/new-technology/1329860/a-brief-hallucination-by-generative-ai-can-land-you-in-hot-water> accessed 22 May 2025

11. Gentile G, 'Trial by Artificial Intelligence: How Technology Is Reshaping Our Legal System' (*LSE - London School of Economics and Political Science*, 8 September 2023) <https://blogs.lse.ac.uk/politicsandpolicy/trial-by-artificial-intelligence-how-technology-is-reshaping-our-legal-system/> accessed 13 May 2025

12. Guha N, Henderson P and Zambrano DA, 'Gamesmanship in Modern Discovery Tech' in Engstrom DF (ed), *Legal Tech and the Future of Civil Justice* (CUP 2023) 112. doi:10.1017/9781009255301.008

13. Güneş Peschke ZS and Peschke L, 'Artificial Intelligence and the New Challenges for EU Legislation' (2022) 2 Yıldırım Beyazıt Hukuk Dergisi 1267. doi:10.33432/ybuhukuk.1104344

14. Kewley J and others, 'Fast Law: Why Speed is the Priority for Lawyers Using AI' (*LexisNexis*, 2024) <www.lexisnexis.co.uk/insights/fast-law-why-speed-is-the-priority-for-lawyers-using-ai/index.html> accessed 2 June 2025

15. Khanna B, *Predictive Justice: Using AI for Justice* (Centre for Public Policy Research 2021)

16. Martin JGC, 'How Far has Technology Invaded the Criminal Justice System?' (ANZELA Legal Studies Teachers' Conference, Brisbane, 11 May 2018)

17. Nowotko PM, 'AI in Judicial Application of Law and the Right to a Court' (2021) 192 Procedia Computer Science 2220. doi:10.1016/j.procs.2021.08.235

18. Pasquale F and Cashwell G, 'Prediction, Persuasion, and the Jurisprudence of Behaviourism' (2018) 68(1) University of Toronto Law Journal 63. doi:10.3138/utlj.2017-0056

19. Rosberger J, 'AI Mediation for Reducing Court Congestion' (26 November 2024) Cornell Journal of Law and Public Policy <https://publications.lawschool.cornell.edu/jlpp/2024/11/26/ai-mediation-for-reducing-court-congestion/> accessed 22 May 2025

20. Sales L, 'Algorithms, Artificial Intelligence and the Law' (2020) 25(1) Judicial Review 46. doi:10.1080/10854681.2020.1732737

21. Skeem JL and Lowenkamp CT, 'Risk, Race, and Recidivism: Predictive Bias and Disparate Impact' (2016) 54(4) Criminology 680. doi:10.1111/1745-9125.12123

22. Socol de la Osa DU and Remolina N, 'Artificial Intelligence at the Bench: Legal and Ethical Challenges of Informing—or Misinforming—Judicial Decision-Making through Generative AI' (2024) 6 Data & Policy e59. doi:10.1017/dap.2024.53

23. Sourdin T and others, 'COVID-19, Technology and Family Dispute Resolution' (2020) 30(4) Australasian Dispute Resolution Journal 270.

24. Surden H, 'Artificial Intelligence and Law: An Overview of Recent Technological Changes in Large Language Models and Law' (2025) 96 University of Colorado Law Review 376. doi:10.2139/ssrn.5135305

25. Surden H, 'The Ethics of AI in Law: Basic Questions' in Dubber MD, Pasquale F and Das S (eds), *The Oxford Handbook of Ethics of AI* (online edn, Oxford Academic 2020) 719. doi:10.1093/oxfordhb/9780190067397.013.46

26. Susskind R, *Online Courts and the Future of Justice* (OUP2019)

27. Taylor I, 'Justice by Algorithm: The Limits of AI in Criminal Sentencing' (2023) 42(3) Criminal Justice Ethics 193. doi:10.1080/0731129X.2023.2275967

28. Winter CK, 'The Challenges of Artificial Judicial Decision-Making for Liberal Democracy' in Bystranowski P, Janik B and Próchnicki M (eds), *Judicial Decision-Making: Economic Analysis of Law in European Legal Scholarship* (Springer Cham 2022) 179. doi:10.1007/978-3-031-11744-2_9

29. Zalnieriute M and Bell F, 'Technology and Judicial Role' in Appleby G and Lynch A (eds), *The Judge, the Judiciary and the Court: Individual, Collegial and Institutional Judicial Dynamics in Australia* (CUP 2021) 116. doi:10.1017/9781108859332

30. Završnik A, 'Criminal Justice, Artificial Intelligence Systems, and Human Rights' (2020) 20 ERA Forum 567. doi:10.1007/s12027-020-00602-0

31. Zhang C and Meng Y, 'Bridging the Divide: Technical Research and Application on Legal Judgment Prediction' [2025] Artificial Intelligence and Law. doi:10.1007/s10506-025-09473-7

AUTHORS INFORMATION

Muhammad Qadeer Ashraf
MA Law (Cont.), European Humanities University, Vilnius, Lithuania
muhammadqadeer.ashraf20@gmail.com
https://orcid.org/0009-0006-8987-1268
Corresponding author, solely responsible for the manuscript preparing.

Competing interests: No competing interests were disclosed.

Disclaimer: The author declares that his opinion and views expressed in this manuscript are free of any impact of any organizations.

RIGHTS AND PERMISSIONS

EDITORS

ABOUT THIS ARTICLE

Cite this article
Ashraf MQ, 'Artificial Intelligence in Courts and Dispute Resolution: Challenges and Opportunities' (2025) 8(Spec) Access to Justice in Eastern Europe 98-118 <https://doi.org/10.33327/AJEE-18-8.S-a000152>

DOI: https://doi.org/10.33327/AJEE-18-8.S-a000152

Keywords: *AI in judicial decision-making, algorithmic fairness in courts, AI governance, human rights and AI regulation, AI and access to justice, AI and legal systems.*

DETAILS FOR PUBLICATION

Date of submission: 15 Jul 2025
Date of acceptance: 05 Oct 2025
Online First publication: 04 Dec 2025
Last Published: 30 Dec 2025
Whether the manuscript was fast tracked? - No
Number of reviewer reports submitted in the first round: 3 reports (2 external reviewers and 1 guest editor)
Number of revision rounds: 1 round with minor revisions

Technical tools were used in the editorial process:
Plagiarism checks - Turnitin from iThenticate https://www.turnitin.com/products/ithenticate/
Scholastica for Peer Review https://scholasticahq.com/law-reviews

AI DISCLOSURE STATEMENT

The author confirmed that AI technologies have only been used to enhance language clarity and grammar. No AI tools were used to generate ideas, structure arguments, analyse data, or produce conclusions.

АНОТАЦІЯ УКРАЇНСЬКОЮ МОВОЮ

Дослідницька стаття

ШТУЧНИЙ ІНТЕЛЕКТ У СУДАХ ТА ВИРІШЕННІ СПОРІВ: ВИКЛИКИ ТА МОЖЛИВОСТІ

Мугаммад Кадір Ашраф

АНОТАЦІЯ

Вступ. *Чи можна використовувати ШІ у судовому процесі без шкоди для правосуддя? Хоча багато хто підтримує потенціал ШІ для підвищення ефективності судової системи, залишаються побоювання щодо його використання в автономному ухваленні рішень та щодо ризиків, які це створює для основоположних прав. Це дослідження оцінює трансформаційний потенціал штучного інтелекту в судовому процесі та аналізує правові проблеми, пов'язані з використанням ШІ в різних сферах, зокрема його допоміжну роль, процесуальні функції та більш суперечливе повністю автономне судочинство. У ньому також досліджується наявна правова база та регуляторні заходи, особливо увагу було звернено на Акт ЄС про ШІ, щоб оцінити, як можна регулювати ШІ, задля забезпечення балансу між ефективністю та захистом прав людини і верховенством права.*

Методи. *У цій статті було використано доктринальний підхід у методології дослідження, проведено ретельний аналіз першоджерел, зокрема Акту ЄС про штучний інтелект, Європейської конвенції з прав людини (ЄКПЛ) та судової практики в юрисдикціях США та Китаю. Крім того, у дослідженні було використано порівняльно-правовий аналіз для вивчення подібностей та відмінностей у регулюванні ШІ, судовій практиці та захисті прав людини в цих юрисдикціях. Було застосовано системний правовий метод під час аналізу Акту ЄС про ШІ у поєднанні з положеннями ЄКПЛ та відповідною судовою практикою для оцінки ширшої правової та регуляторної бази, що застосовується до ШІ у судовому процесі. Вторинні джерела, такі як статті в наукових журналах, книги, дисертації, доповіді на конференціях, газетні статті, звіти та блоги, були критично оцінені за допомогою тесту CRAAP. Нарешті, було застосовано критичний правовий метод для оцінки етичних, процедурних та правозахисних наслідків впровадження ШІ у судовому процесі.*

Результати та висновки. *Результати цього дослідження підтверджують використання ШІ як допоміжної функції в судовому процесі, а не як повністю автономного органу, що ухвалює рішення. Судочинство за допомогою ШІ може бути посилено завдяки надійним етичним та процесуальним гарантіям, зокрема обов'язковому людському контролю, підвищеній прозорості та покращенню інтерпретації алгоритмічних рішень. Такі заходи можуть зменшити ризики, пов'язані зі справедливістю, упередженістю та руйнуванням судової дискреції, виявлені в повністю автономних системах. Коли інструменти штучного інтелекту функціонують виключно як допоміжні, а не як авторитетні суб'єкти, занепокоєння*

щодо порушень права на справедливий суд, незалежність судової влади та ефективну участь значно знижуються. Зрештою, легітимність ШІ в ухваленні рішень у суді залежить не стільки від його самої присутності, скільки від того, як він розроблений, регулюється та сприймається як судовою системою, так і громадськістю. Ґрунтуючись на цих висновках, це дослідження пропонує конкретні рекомендації для судових установ, щоб забезпечити підвищення ефективності ШІ без шкоди для правосуддя.

Ключові слова. *ШІ в ухваленні судових рішень, алгоритмічне правосуддя у судах, управління ШІ, права людини та регулювання ШІ, ШІ та доступ до правосуддя, ШІ та правові системи.*

ABSTRACT IN ARABIC

مقال بحثي

الذكاء الاصطناعي في المحاكم وتسوية المنازعات: التحديات والفرص

محمد قدير أشرف

الملخص

خلفية الدراسة: *هل يمكن توظيف الذكاء الاصطناعي في عملية الفصل القضائي دون الإضرار بمبدأ العدالة؟ في الوقت الذي يؤيد فيه كثيرون قدرة الذكاء الاصطناعي على تحسين كفاءة العمل القضائي، ما تزال المخاوف قائمة بشأن استخدامه في اتخاذ القرارات بشكل مستقل وما قد يترتب على ذلك من مخاطر تمس الحقوق الأساسية. يقيّم هذا البحث الإمكانات التحولية للذكاء الاصطناعي في القطاع القضائي، ويحلل التحديات القانونية المرتبطة باستخدامه في أدوار متعددة، تشمل المساعدة الإجرائية والوظائف التقنية، إضافة إلى الدور الأكثر جدلاً المتعلق بالفصل القضائي الذاتي بالكامل. كما يستعرض البحث الأطر القانونية القائمة والاستجابات التنظيمية، مع تركيز خاص على قانون الذكاء الاصطناعي الأوروبي، بهدف تقييم كيفية تنظيم هذه التقنيات بما يحقق التوازن بين الكفاءة من جهة، وحماية حقوق الإنسان وسيادة القانون من جهة أخرى.*

المنهجية: *اعتمد هذا البحث منهجًا فقهيًا يقوم على تحليل شامل للمصادر الأساسية، بما في ذلك قانون الاتحاد الأوروبي للذكاء الاصطناعي، والاتفاقية الأوروبية لحقوق الإنسان، والأحكام القضائية الصادرة في الولايات المتحدة والصين. إضافة إلى ذلك، استخدمت الدراسة منهج المقارنة القانونية لتحليل أوجه التشابه والاختلاف في تنظيم الذكاء الاصطناعي والممارسات القضائية وضمانات حقوق الإنسان عبر*

هذه النظم القانونية. وقد تم تطبيق منهج قانوني منهجي من خلال تحليل قانون الذكاء الاصطناعي الأوروبي جنبًا إلى جنب مع أحكام الاتفاقية الأوروبية لحقوق الإنسان والاجتهادات القضائية ذات الصلة بهدف تقييم الإطار القانوني والتنظيمي الأشمل الذي يحكم استخدام الذكاء الاصطناعي في عملية الفصل القضائي. كما جرى تقييم المصادر الثانوية، بما في ذلك المقالات العلمية والكتب والرسائل الجامعية وأوراق المؤتمرات والمقالات الصحفية والتقارير والمدونات، وفق اختبار CRAAP. وفي النهاية، طُبّق منهج قانوني نقدي لتقدير الأبعاد الأخلاقية والإجرائية وتأثيرات الذكاء الاصطناعي على حقوق الإنسان عند توظيفه في صنع القرار القضائي.

النتائج والاستنتاجات: تؤكد نتائج هذا البحث جدوى توظيف الذكاء الاصطناعي في دور مساعد ضمن عملية الفصل القضائي، بدلاً من الاعتماد عليه كصانع قرار مستقل بشكل كامل. ويمكن تعزيز فعالية هذا الدور المساعد من خلال وضع ضمانات أخلاقية وإجرائية قوية تشمل الإلزام بالرقابة البشرية، ورفع مستويات الشفافية، وتحسين قابلية تفسير القرارات الخوارزمية. تسهم هذه التدابير في الحد من المخاطر المتعلقة بالإنصاف والتحيز وتراجع السلطة التقديرية للقاضي، وهي المخاطر التي ترتبط بالأنظمة القائمة على الاستقلالية الكاملة. وعندما تعمل أدوات الذكاء الاصطناعي بوضوح كوسائل دعم لا كجهات ذات سلطة تقريرية، تنخفض بدرجة كبيرة المخاوف المرتبطة بانتهاك الحق في محاكمة عادلة واستقلال القضاء وفعالية مشاركة الأطراف. وفي نهاية المطاف، تعتمد شرعية استخدام الذكاء الاصطناعي في صنع القرار داخل قاعات المحاكم على كيفية تصميمه وتنظيمه وصورة استخدامه لدى السلطة القضائية والجمهور، أكثر مما تعتمد على مجرد وجوده. واستنادًا إلى هذه النتائج، يقدم هذا البحث توصيات عملية للمؤسسات القضائية لضمان أن يسهم الذكاء الاصطناعي في تعزيز الكفاءة دون المساس بالعدالة.

Access to Justice in Eastern Europe
ISSN 2663-0575 (Print)
ISSN 2663-0583 (Online)
http://ajee-journal.com

Research Article

ARTIFICIAL INTELLIGENCE AND LAW: PROCEDURAL SAFEGUARDS AND REGULATORY CHALLENGES IN KAZAKHSTAN

Anuar Nurmagambetov*, Anet Nurmagambetov, Amanzhol Nurmagambetov and Aigerim Zhumabayeva

ABSTRACT

***Background**: The active integration of artificial intelligence (AI) into diverse spheres of human activity has created significant opportunities for innovation and efficiency, while simultaneously raising complex ethical, legal, and social challenges. Among these, the deployment of high-risk AI systems requires particular scrutiny due to their potential impact on fundamental rights, public safety, and socio-economic relations. This research examines both the benefits and risks of AI technologies, with an emphasis on the need to establish clear legal and regulatory frameworks at the national and international levels.*

***Methods**: The study employs a comparative legal analysis of existing regulatory approaches, including the European Union's AI Act (EU AI Act), the OECD AI Principles, and national legislative practices. The methodology is based on a systematic review of normative legal acts, doctrinal sources, and policy papers, as well as an evaluation of prospective risks associated with the use of high-risk AI systems in various sectors, including transport, healthcare, and financial services.*

DOI:
https://doi.org/10.33327/AJEE-18-8.S-a000157

Date of submission: 25 Aug 2025
Date of acceptance: 28 Oct 2025
Online First publication: 04 Dec 2025
Last Published: 30 Dec 2025

Disclaimer:

The authors declare that their opinion and views expressed in this manuscript are free of any impact of any organizations.

***Results and conclusions**: The analysis reveals that, while the adoption of AI contributes to economic development, efficiency in public administration, and improved quality of services, it also generates risks such as discrimination, violations of privacy, cyberthreats, and reduced accountability. In particular, the study highlights that existing legislation in Kazakhstan, as in many other jurisdictions, does not sufficiently address the specificities of high-risk AI systems. Comparative legal analysis demonstrates that the most effective regulatory models are risk-oriented, ensuring transparency, human oversight, and liability mechanisms. The findings suggest that partial amendments to existing legislation—such as in the areas of mandatory insurance and consumer protection—could serve as an interim measure, while the adoption of a dedicated AI law may be necessary in the long term.*

The study underscores the need for a balanced legal framework that harmonises technological innovation with the protection of human rights and societal interests. It is argued that Kazakhstan, while considering international best practices, should pursue a two-stage approach: (1) introducing targeted amendments to sectoral legislation; and (2) elaborating a comprehensive AI law focused on high-risk systems. Such a framework would mitigate risks, ensure accountability, and foster public trust, while promoting the responsible and sustainable use of artificial intelligence.

1 INTRODUCTION

Alan Turing's extensive research into artificial intelligence (AI) laid the groundwork for empirical methods to evaluate the capabilities of early computers in the late 1940s. However, the term "artificial intelligence" was coined later, emerging as the subject of a university course at Dartmouth College in 1956.[1] Theorising, testing, implementing, optimising and regulating AI processes and applications attracts an increasing number of experts from various fields, including law. Turing aptly noted, "There are signs ... that it is possible to make a machine demonstrate intelligence, while at the risk of making serious mistakes from time to time ... The whole process of thinking is still rather mysterious to us, but I believe that the attempt to create a thinking machine will greatly help us in finding out how we think ourselves."[2]

One of the priority areas of development of the Kazakh economy in the near future is the creation of an AI technology industry. President Kassym-Jomart Tokayev has stated, "We need to turn our country into a place of attraction for 'digital nomads' from all over the

1 'Artificial Intelligence Coined at Dartmouth, 1956' (*Dartmouth College*, 2025) <https://home.dartmouth.edu/about/artificial-intelligence-ai-coined-dartmouth> accessed 20 August 2025.

2 B Jack Copeland, 'History of Artificial Intelligence (AI): Alan Turing and the Beginning of AI', *Britannica* (7 November 2025) <https://www.britannica.com/science/history-of-artificial-intelligence> accessed 18 November 2025.

world". The President claims that, "Our success in all other areas depends on how quickly and effectively we develop new digital technologies."[3]

In response, the Government of the Republic of Kazakhstan is currently developing a Strategy for the Development of Artificial Intelligence, a Digital Code and the Law "On Artificial Intelligence". The Concept for the Development of AI for 2024–2029, approved by the Resolution of the Government of the Republic of Kazakhstan dated 24 July 2024, No. 592, outlines the current state of AI, assesses preliminary readiness, evaluates the research base, reviews international experience and establishes the basic principles and approaches for AI development.[4]

Kazakhstani legal research increasingly focuses on defining AI, its legal capacity, liability, and ethical standards amid insufficient regulation.[5] Defining the key challenges associated with the implementation and operation of AI is intended to facilitate the development of a comprehensive legal model, unify approaches to problem-solving and, ultimately, increase the effectiveness of legal regulation.[6]

In his Address to the People of Kazakhstan, President Tokayev emphasised that, "to become part of the new technological paradigm, it will be necessary to restructure the entire system of public administration with a manifold increase in its transparency, efficiency, and human-centred orientation".[7] This underscores a critical issue: the legal regulation of AI integration into public administration and its interaction with the protection of citizens' procedural rights. This challenge is both ethical and regulatory, representing a fundamental aspect of the relationship between society and the state in the era of digital development.

3 Alexandra Golm, 'Tokayev Spoke about the Development of AI and the Creation of a Supercomputer in Kazakhstan' (*NUR.KZ*, 12 April 2024) <https://www.nur.kz/technologies/software/2083760-tokaev-vyskazalsya-o-razvitii-ii-i-sozdanii-superkompyutera-v-kazahstane/> accessed 20 August 2025.

4 Resolution of the Government of the Republic of Kazakhstan No 592 'On approval of the Concept for the Development of Artificial Intelligence for 2024–2029' (24 July 2024) <https://adilet.zan.kz/kaz/docs/P2400000592> accessed 20 August 2025.

5 Zhanna U Tlembayeva, 'On Some Approaches to the Legal Regulation of Artificial Intelligence' (2021) 2(65) Bulletin of the Institute of Legislation and Legal Information of the Republic of Kazakhstan 61. doi:10.52026/2788-5291_2021_65_2_61.

6 Zhanna U Tlembayeva, 'On Legal Regulation of the Use of Artificial Intelligence in Healthcare of the Republic of Kazakhstan' (2022) 5-1 Greater Eurasia: Development, Security, Cooperation 1123; Darya Zhanysbayeva and Mila Ryzhkina, 'Draft Law of the Republic of Kazakhstan "On Artificial Intelligence" – Principles of Regulation and Practical Aspects' (*GRATA International*, 14 July 2025) <https://gratanet.com/publications/draft-law-of-the-republic-of-kazakhstan-on-artificial-intelligence-principles-of-regulation-and-practical-aspects > accessed 9 September 2025.

7 Kassym-Jomart Tokayev, 'Kazakhstan in the Era of Artificial Intelligence: Current Challenges and Solutions through Digital Transformation: President's State of the Nation Address to the People of Kazakhstan' (*Әділет*, 8 September 2025) <https://adilet.zan.kz/kaz/docs/K25002025_1> accessed 10 November 2025.

2 METHODOLOGY

This study examines the legal regulation of AI in Kazakhstan is reviewed in this study using a comprehensive approach that integrates both theoretical analysis and practical insights into existing developments and practical recommendations.

The study's basis is the comparative legal method, which enables a comparative analysis of legal practices and regulations across countries. This method was applied to analyse relevant texts and policies from the European Union, the United States of America, China, and the Republic of Kazakhstan.[8] While the EU, US, and Chinese frameworks have already been in use for quite some time, Kazakhstan has only recently adopted the Law "On Artificial Intelligence".[9] The comparative legal method allows for the identification of global trends and country-specific features, guiding the adaptation of these developments to the Republic of Kazakhstan's context. For instance, the EU's *AI Act* illustrates how to balance AI-driven economic innovation with risk reduction,[10] while the *White Paper on AI Development*, developed in China, provides a wealth of information on the introduction of AI across all areas of society.[11]

To understand the nature of AI itself, the study employs a systems analysis, conceptualising AI as a full-fledged system in which each component is responsible for the functioning of the others; that is, it is not an isolated system but a well-coordinated one. In this regard, it is possible to identify economic, social, technological, cultural and other aspects of its existence. Conducting such an analysis is extremely necessary for adapting AI tools to traditional sectors of Kazakhstan's economy, using new tools to improve efficiency without destroying these industries.[12]

Given widespread public concerns about AI's impact, ethical considerations are a key aspect of AI deployment, and the ethical dimension is an essential component of this

8 ibid

9 Law of the Republic of Kazakhstan No 230-VIII "On Artificial Intelligence" (17 November 2025) <https://online.zakon.kz/Document/?doc_id=33005677> accessed 18 November 2025.

10 Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) [2024] OJ L 1689 <http://data.europa.eu/eli/reg/2024/1689/oj> accessed 20 August 2025; 'The EU AI Act: Up-to-date Developments and Analyzes of the EU AI Act' (*EU Artificial Intelligence Act*, 2025) <https://artificialintelligenceact.eu/> accessed 20 August 2025.

11 CAICT, *Artificial Intelligence White Paper (2022)* (*CSET Center for Security and Emerging Technology within Georgetown University's Walsh School*, 16 June 2022) <https://cset.georgetown.edu/publication/artificial-intelligence-white-paper-2022/> accessed 20 August 2025.

12 Svetlana Moroz and Saparmurat Muzaparov, 'Problems of Copyright and Intellectual Property Rights in Connection with the Use of AI Technologies (Neural Networks) (2024) 198 Scientific Collection InterConf 260 <https://archive.interconf.center/index.php/conference-proceeding/article/view/6025> accessed 20 August 2025.

study. AI has generated numerous discussions on ethical issues such as legal capacity, who should be held responsible for the damage caused by AI, and how appropriate it is to use it in those spheres of society where direct human participation has traditionally been required—such as education (e.g., writing qualification papers) and medicine (e.g., diagnostics and postoperative care recommendations). This method was used to analyse recommendations from the OECD[13] and the UN Global Digital Compact.[14] Based on this analysis, proposals were formulated to integrate AI into Kazakhstan's legislation, while accounting for potential ethical issues.

Using the method of critical analysis of regulatory documents, the Law "On Artificial Intelligence" and the Concept of AI Development for 2024-2029 were examined. The results revealed gaps in the legislation, particularly regarding the definition of AI, its legal status, and potential liability in the event of damage.[15]

The three proposed regulatory models were derived from a comparative analysis of international legal frameworks and adapted to Kazakhstan's institutional and socio-economic context, as further detailed in the Results and Discussion section (Table 4).

In assessing AI, particular attention was paid to the potential risks associated with its practical application. A meaningful discussion of these risks requires the modelling method. Based on the study of international experience and current trends in Kazakhstan, three models of legal regulation of AI regulation were formulated: (1) copying international experience, (2) a symbiosis of international and national approaches, and (3) minimal regulation to gain a technological advantage.[16]

Thus, the chosen methodological framework enabled comprehensive coverage of the identified problem, identification of existing shortcomings in the legislation, and determination of the key directions for its improvement through specific paths and measures tailored to Kazakhstan's context.

13 OECD, 'Recommendation of the Council on Artificial Intelligence' (*OECD Legal Instruments*, 22 May 2019) <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449> accessed 10 April 2025; Elvira S Kuandykova, Daulet L Baideldinov and Thomas Hoffmann, 'Problems of Legal Regulation of Digital Transformation of Agriculture of the Republic of Kazakhstan' (2023) 112(4) Bulletin of the Karaganda University: Law Series 7. doi:10.31489/2023L4/7-17.

14 UNGA Resolution A/RES/79/1 'The Pact for the Future: Annex I Global Digital Compact' (22 September 2024) <https://docs.un.org/en/a/res/79/1> accessed 20 August 2025.

15 Maidan K Suleimenov and Farkhad S Karagusov, "The Concept of Recodification of the Civil Code of Ukraine and Modernization of the Civil Code of Kazakhstan: A Comparative Analysis of the Main Ideas" (*Paragraph Lawyer*, 29 July 2021) <https://online.zakon.kz/Document/?doc_id=32892885> accessed 20 August 2025.

16 Ricardo Francisco Reier Forradellas and Luis Miguel Garay Gallastegui, "Digital Transformation and Artificial Intelligence Applied to Business: Legal Regulations, Economic Impact and Perspective' (2021) 10(3) Laws 70. doi:10.3390/laws10030070.

The selection of normative and policy sources was based on official legislative databases and international repositories, including Adilet (Kazakhstan), ISO/IEC standards archives, the UN Global Digital Compact, and the EU AI Act documentation. Only publicly accessible and officially adopted acts, standards, and policy papers were included in the analysis.

This study is limited to the analysis of normative and policy documents and does not include empirical data on enforcement or judicial practice. Future research could extend the analysis to case law and administrative decisions to assess how AI-related norms are applied in practice.

3 RESULTS AND DISCUSSION

The legal issues surrounding AI are extensive and warrant thorough theoretical research. The focus of this article will be on some of the legal issues of theorising and practical implementation of legislation on AI in light of the adoption of the Law "On Artificial Intelligence" (hereinafter: the Law "On Artificial Intelligence") by the Ministry of Digital Development, Innovation and Aerospace Industry of the Republic of Kazakhstan.[17] Developing coherent AI legislation is crucial for Kazakhstan's theory, law-making, and practice.[18]

3.1. Ethical and Copyright Challenges

3.1.1. Problems of Legal Regulation of Artificial Intelligence

The Concept of AI Development in Kazakhstan for 2024-2029 (hereinafter: the Concept) is a significant step towards the introduction of AI into the economy and society of Kazakhstan.[19] Based on the Concept data, Kazakhstan's readiness for AI is reflected in the 45.78 indicator. The country ranks 72nd out of 181 countries on this list. The most "prepared" categories for working with AI are "digital potential" (75.67), "adaptability" (63.76), and "data availability" (74.11). At the same time, there are serious problems associated with society's unpreparedness for these changes. Thus, in particular, the category "infrastructure" is estimated at only 30.80, and "human capital" at 38.55.[20] This indicates that the country has few qualified personnel ready to work with artificial intelligence, and a

17 Law of the Republic of Kazakhstan No 230-VIII (n 9); Tlembayeva, 'On Legal Regulation of the Use of Artificial Intelligence in Healthcare of the Republic of Kazakhstan' (n 6).

18 Resolution of the Government of the Republic of Kazakhstan No 592 (n 4).

19 Sarah K Idrysheva, 'On the Digital Code of Kazakhstan' (2022) 3(96) Law and State 72. doi:10.51634/2307-5201_2022_3_72.

20 Mohamed Hamada and others, 'Artificial Intelligence to Improve the Business Efficiency and Effectiveness for Enterprises in Kazakhstan' (2021) 4(1) SAR Journal - Science and Research 34. doi:10.18421/SAR41-06.

sufficient material and technical base for this has not been created. The indicator "vision (strategy)" is 0, and "maturity", that is, readiness for AI, is 15.48. As a result, the introduction of AI poses a major problem across all areas of Kazakhstani society.

Based on the above, the Concept emphasises the need to create conditions for working with AI and for sustainable growth in this area, which will require major changes and additions in the legislative sphere. One of the main problems is the uncertainty of existing definitions, as well as legal mechanisms regulating AI and responsibility for its use.[21] The Concept appropriately notes that the existing trends in informatisation and cybersecurity development in Kazakhstan do not meet the requirements for addressing problems related to the use of artificial intelligence.

The developers of the Law "On Artificial Intelligence" understand AI as "the functional ability to imitate cognitive functions characteristic of humans, providing results comparable to or superior to those of human intellectual activity".[22] The difficulties of applying the concept of "artificial intelligence" will most likely 'haunt' the Kazakh legislator, as it is currently quite difficult to determine with a high degree of accuracy the range of issues associated with the development of such a broad phenomenon. It is characteristic that the recently adopted EU AI Act provides an operational definition of AI (Article 3), focusing on the system's functionality and associated risks rather than a single conceptual definition. This cautious approach reflects the EU's intent to regulate AI based on risk categories.[23]

***Table* 1. Risk-based grading of AI systems in the EU**

AI System Class	Measures and requirements applied	Examples
Prohibited practices	Complete ban on the use of malicious AI systems.	Systems for inciting violence.
High risk	Mandatory registration, security requirements, cybersecurity, data management.	Biometrics, autonomous vehicles.
Limited risk	Compliance with the principles of process transparency, minimum requirements.	Recommender systems in trading.
Low risk	No obligations or restrictions.	Virtual assistants, simple chatbots.

The OECD has taken a distinctive approach to defining artificial intelligence, revising its definition in 2023 to describe AI as "systems whose behaviour can be haracterized as intelligent. This includes the ability to learn from data, adapt to new inputs, and perform

21 Tlembayeva, 'On Some Approaches to the Legal Regulation of Artificial Intelligence' (n 5).

22 Law of the Republic of Kazakhstan No 230-VIII (n 9).

23 Idrysheva (n 19).

tasks that would normally require human intelligence."[24] The OECD deliberately uses such a broad formulation to encompass various AI technologies, including both machine learning and rule-based systems. However, the OECD also defines an AI system as "a computer system that, in order to achieve explicit or implicit goals, determines from the inputs it receives how to generate outputs such as predictions, content, recommendations or decisions that can influence a physical or virtual environment".[25]

In the broader context of developing international standards for the legal regulation of AI, most researchers consider it necessary to be guided by the requirements of AI system security and transparency. Thus, the ISO/IEC 38507:2022 and ISO/IEC 23894:2023 standards note that, first, a basis is needed to ensure cybersecurity and prevent AI-related fraud and crime.[26]

***Table* 2. Key requirements for high-risk AI systems in the EU**

Requirement	Description
Registration in the database	Developers are required to register AI before entering the market.
Risk management	Mandatory testing for technical reliability, training and data management.
Transparency	Developers are required to disclose information about the algorithms they use and how they train models.
Cybersecurity	Strict requirements for ensuring the security of AI systems.
Post-Market Monitoring	Post-marketing monitoring by regulators.

Different AI systems vary in their level of autonomy and adaptability after deployment. In this sense, the developers of the Laws "On Artificial Intelligence" plan to distinguish AI from AI systems. They propose the following definition: "AI technologies (systems) - technologies based on the use of artificial intelligence, including speech and visual image recognition, analytical decision-making, complex logical operations, and intelligent decision support."[27]

24 Kuandykova, Baideldinov and Hoffmann (n 13).

25 Idrysheva (n 19).

26 ISO/IEC 38507:2022 Information Technology - Governance of IT - Governance Implications of the Use of Artificial Intelligence by Organizations (2022) <https://www.iso.org/standard/56641.html> accessed 20 August 2025; ISO/IEC 23894:2023 Information Technology - Artificial Intelligence - Guidance on Risk Management (2023) <https://www.iso.org/standard/77304.html> accessed 20 August 2025.

27 Patricia Gomes Rêgo de Almeida, Carlos Denner dos Santos and Josivania Silva Farias, 'Artificial Intelligence Regulation: A Framework for Governance' (2021) 23 Ethics and Information Technology 505. doi:10.1007/s10676-021-09593-z.

***Table* 3. Comparison of international approaches to AI regulation**

Country	Definition of AI	Degree of regulation	Ethical standards	Problems
EU	Does not contain a clear definition of AI	Moderate, AI Act	Transparency, control	Safety, responsibility
Kazakhstan	Information and communication technology	Law "On Artificial Intelligence"	Ethics, control	Lack of theoretical and practical basis
USA	Does not contain a clear definition of AI	Liberating regulation	Flexibility, ethics	Predictability of AI behavior
China	Does not contain a clear definition of AI	Strict standards and control	Restrictions in certain areas	Ethical issues

If the technological aspect of defining the concept of artificial intelligence is not considered at this stage, it is proposed that the Kazakh legislator should refrain from distinguishing between AI and AI systems (including rule-based systems, as understood by the OECD). Given the complexity of the phenomenon and the lack of clear criteria for determining the "intelligence" of emerging systems, such an approach may lead to further difficulties in classifying particular technologies as AI systems. At present, legal science in the field of AI faces three main problems of defining the concept, nature and limits of regulation:

1. Semi-autonomy: A certain degree of autonomy may lead to unexpected results, and therefore, the creation of systems that adapt to the changing nature of AI will be difficult due to the over-regulation of developments in this area.[28] Researchers note that, in the case of increasing the autonomy of AI, it is important to correctly balance between the freedom of algorithms and the need for legislative control.[29]

2. Predictability: The increasing complexity of AI systems reduces the predictability of their behaviour, which, to a certain extent, levels out the possibilities of management and regulation at the regulatory level.[30] In this context, the importance of creating flexible legal systems that can adapt to rapid technological changes is discussed.[31]

3. Unlimited application: The variability of the application of AI in various areas of human activity complicates the possible regulatory structure, and the construction of a system of norms regulating various areas of application of AI may be a massive undertaking that does

28 Hamada and others (n 20).
29 de Almeida, dos Santos and Farias (n 27).
30 Kuandykova, Baideldinov and Hoffmann (n 13).
31 Regulation (EU) 2024/1689 (n 10).

not keep pace with the development of relevant technologies.[32] Thus, it is necessary to develop mechanisms that enable rapid, effective regulation of new AI uses across sectors.[33]

The abovementioned problems, in a broad sense, prevent the determination of the legal capacity of artificial intelligence and preclude its recognition as a full-fledged subject of law. In this regard, Suleimenov and Karagusov reasonably argue that "Neither robots nor AI should be recognised as subjects of law. At the level of the Civil Code, their legal regime should be enshrined as a separate category of objects of civil rights, excluding human interaction with them, allowing only human influence on them."[34]

The lack of a clear legal status for AI underscores the need to define appropriate regulatory frameworks for its use. To address this issue, three alternative regulatory models were developed based on comparative legal analysis, as summarised below.

***Table* 4. Comparative characteristics of AI regulatory models**

Model	Description	Advantages	Limitations	Applicability to Kazakhstan
Direct adoption of international norms	Copying EU and OECD standards and regulatory practices	Legal harmonisation, predictability, compliance with global norms	Low adaptability to national context, potential overregulation	Moderate
Hybrid (international + national)	Combination of global standards with local legal and institutional specifics	Flexibility, contextual relevance, balanced regulation	Requires strong institutional capacity and policy coherence	High
Minimal regulation	Limited legal interference to foster innovation	Technological advantage, fast implementation	Legal uncertainty, weak protection of rights	Selective / Experimental

The comparative analysis demonstrates that each regulatory model offers distinct advantages and challenges for the legal governance of AI in Kazakhstan. The first model, direct adoption of international norms, provides the most predictable and harmonised

32 ISO/IEC 38507:2022 (n 26).

33 Keng Siau and Weiyu Wang, 'Artificial Intelligence (AI) Ethics: Ethics of AI and Ethical AI' (2020) 31(2) Journal of Database Management 74. doi:10.4018/JDM.2020040105.

34 Suleimenov and Karagusov (n 15).

approach. By aligning national legislation with the EU AI Act and OECD standards, this model would ensure compliance with global norms, facilitate international cooperation, and simplify cross-border data and technology exchange. However, its primary drawback lies in the limited adaptability of international standards to Kazakhstan's socio-economic and legal environment. Excessive reliance on external regulatory templates could lead to overregulation and hinder local innovation.

The hybrid model, combining international standards with national regulatory mechanisms, is the most balanced and contextually suitable option. It allows Kazakhstan to maintain consistency with international best practices while tailoring specific provisions to national realities. Such an approach supports flexibility, promotes institutional learning, and enables gradual adaptation of legal norms as technologies evolve. Nevertheless, its effective implementation requires a high degree of institutional coordination, capacity building, and sustained policy coherence—areas that currently remain underdeveloped.

The third model—minimal regulation—represents a liberal framework aimed at stimulating innovation and rapid technological development. It provides significant room for experimentation and entrepreneurship in the AI sector. Yet, the absence of clear legal safeguards increases the risks of legal uncertainty, ethical violations, and insufficient protection of human rights. Consequently, this model can be applied only selectively, for instance, in pilot projects or regulatory sandboxes.

Overall, the analysis suggests that the hybrid model offers the most practical pathway for Kazakhstan, as it balances innovation incentives with the need for legal certainty and social accountability, aligning technological progress with national institutional capacity.

3.1.2. Problems of Normative and Technical Regulation of Artificial Intelligence

Unclear legal definitions complicate the technical regulation of AI. Kazakhstan has yet to develop a set of national standards for the technical regulation of artificial intelligence. Still, the legislator should already be considering building a coherent legal framework to minimise potential risks in the creation and use of artificial intelligence, its systems, and robotics based on it. At the same time, the base of international standards for technical regulation of AI is developing very quickly, where the main risks are the following (Table 5):

1. Reputational costs to the owner of AI in the event of harm caused to others due to a lack of control over artificial intelligence;
2. Complete or partial loss of control over the exploited artificial intelligence;
3. Disenfranchisement of workers whose functions are replaced by the work of artificial intelligence;
4. Univariance in the judgments of AI when processing data due to the limitations of the data provided to it;

5. Increased complexity of competition between market participants using AI and those who do not; Difficulty in predicting the performance of AI due to limited historical data and rapidly changing future expectations.[35]

***Table* 5. Problems of normative and technical regulation of AI**

Risk	Description	Proposed measures
Reputational costs	Harm caused to others due to lack of control	Codes of Ethics, Voluntary Standards
Loss of control	Complete or partial loss of control over AI	Transparency, monitoring
Unlimited application	Variability of AI application across industries	Development of industry standards
Difficulties in forecasting	Problems with predictability of AI behavior	Improving AI learning systems

Currently, the difficulties associated with the legal regulation of AI, as well as the predicted risks, are only increasing. The White Paper on the Development of AI in China (April 2024), among the technical risks, points to possible delusions of artificial intelligence, in which systems produce answers or judgments that do not correspond to reality, such as when processing images or language structures.[36] When such systems are used in areas like healthcare or transport, these errors may pose risks to citizens' lives and health. To solve these problems, the document proposes implementing AI learning systems with creator feedback (RLHF) or "Fence technology" (NeMo Guardrails). However, the deployment of these solutions remains challenging. This is why the risks of using AI in medicine and healthcare require primary attention and legal regulation. B. Murdoch, a scientist who described cases of using AI in hospitals in his work, agrees with this. The author states that the lack of legal regulation leads to a violation of confidentiality and unauthorised "replication of medical information."[37]

A risk-oriented approach to national standardisation and regulation of AI in the Republic of Kazakhstan appears justified under current conditions. Building a system of standards should be based on the following categories of risk:

1. Comparable risk: The introduction of AI systems such as AI-enabled video games, spam filtering systems, and other similar technologies should not, for the most part,

35 ISO/IEC 38507:2022 (n 26); ISO/IEC 23894:2023 (n 26); ISO/IEC 22989:2022 Information Technology - Artificial Intelligence - Artificial Intelligence Concepts and Terminology (2022) <https://www.iso.org/standard/74296.html> accessed 20 August 2025.

36 CAICT (n 11).

37 Blake Murdoch, 'Privacy and Artificial Intelligence: Challenges for Protecting Health Information in a New Era' (2021) 22 BMC Med Ethics 122. doi:10.1186/s12910-021-00687-3.

create obligations and liabilities for developers. However, it is necessary to create conditions in which developers voluntarily follow codes of ethics and conduct, thereby minimising potential risks. To this end, it is important to develop mechanisms that encourage companies to self-regulate and implement best practices that help protect users.[38]

2. Transparency risk: AI-based systems must clearly inform consumers when interactions involve a machine, especially in areas such as chatbots and content-generating systems. This will help increase user trust and ensure transparency in interactions with AI systems. At the same time, measures in this area must be balanced with the need to protect personal data and comply with privacy standards.[39]

3. Contact risk: AI systems operating in sensitive areas—such as medicine, military technology, and selection or screening processes—must be subject to strict restrictions that reflect potential risks to human health and safety. Human oversight remains essential to ensure that AI use does not lead to potentially dangerous consequences. Strict regulations aimed at managing these risks must take into account both technical and ethical aspects.[40]

4. Manipulative risk: It is important to embed moral and ethical principles into the foundation of AI technologies to avoid manipulative abuses. For example, aggressive sales, social scoring, big data processing and other types of manipulation can lead to serious consequences for human rights. AI developers must take these risks into account and develop systems that ensure the rights and freedoms of citizens are respected, preventing the misuse of AI to manipulate personal data.[41]

Based on the content of the Law "On Artificial Intelligence",[42] it appears that the Kazakh legislator is taking the path of excessive regulation, in which the authorised body in the field of AI will maintain a classifier of AI systems, which will prohibit the creation, development and operation of systems with capabilities other than those defined by the classifier. The introduction of such a classifier contradicts the principles enshrined in the Concept of AI Development for 2024-2029,[43] which holds that a low level of regulation can provide a technological advantage. In this sense, the quality of the classifier leaves questions, and AI systems, especially dual-use ones, may not be developed.

38 Siau and Wang (n 33).
39 ISO/IEC 38507:2022 (n 26).
40 Hamada and others (n 20).
41 Nicola Lucchi, 'ChatGPT: A Case Study on Copyright Challenges for Generative AI Systems' [2023] European Journal of Risk Regulation 1. doi:10.1017/err.2023.59.
42 Law of the Republic of Kazakhstan No 230-VIII (n 9).
43 Resolution of the Government of the Republic of Kazakhstan No 592 (n 4).

3.1.3. Problems of Ethical Regulation of Artificial Intelligence

Many people are concerned about the ethical issues arising from the use of AI. According to researcher L. Floridi, these issues should be addressed not only through theoretical development but also through implementation in current and/or developing legislation. This will minimise ethical risks, prevent a possible decline in the reputations of those who use AI in their activities, and prevent potential human rights violations associated with these systems.[44] T. Hagendorf agrees with this opinion, according to whom universal principles of ethical regulation of the use of AI should be developed initially, after which all countries will be guided by them when developing legislative norms and using AI in practice.[45]

While these international frameworks provide valuable guidance, their practical application in Kazakhstan requires adaptation to national constitutional principles and legal traditions. Ethical principles such as transparency, fairness, and accountability resonate with Kazakhstan's constitutional provisions, including the right to privacy (Article 18) and the prohibition of discrimination (Article 14).[46] Ethical self-regulation by developers can complement formal legal regulation, ensuring flexibility and innovation within the boundaries of legal accountability. Together, these mechanisms form a balanced framework that aligns ethical responsibility with the rule of law in Kazakhstan.

AI ethics challenges arise from the difficulty of defining ethical principles for its creation and operation. The complete absence of regulatory decisions in this area in Kazakhstan at the initial stage allows reliance on existing international legal developments. Nevertheless, even international practice lacks a unified approach to the formation of a basket of ethical principles for AI and a strict hierarchy for them, which complicates the development of a general ethical basis. Significant progress in AI development and regulatory developments in several countries further complicate the harmonisation of international norms with national legislation, both due to differing levels of technology penetration and the number of ethical and legal documents adopted at the national level.

At the same time, the scientific community mainly concentrates on the study of the ethical principles of the functioning of AI either in relation to the abstract definition of a "good" or "bad" algorithm (rather than "well-designed" and "poorly designed"),[47]

44 Luciano Floridi, 'Introduction to the Special Issues: The Ethics of Artificial Intelligence: Exacerbated Problems, Renewed Problems, Unprecedented Problems' (2024) 61(4) American Philosophical Quarterly 301. doi:10.5406/21521123.61.4.01.

45 Thilo Hagendorff, 'The Ethics of AI Ethics: An Evaluation of Guidelines' (2020) 30 Minds & Machines 99. doi:10.1007/s11023-020-09517-8.

46 Constitution of the Republic of Kazakhstan of 30 August 1995 (amended 18 July 2025) <https://adilet.zan.kz/eng/docs/K950001000_> accessed 20 August 2025.

47 Jessica Morley and others, 'Operationalising AI Ethics: Barriers, Enablers and Next Steps' (2023) 38 AI & Society 411. doi:10.1007/s00146-021-01308-8.

through the lens of specific regulatory domains,[48] or in broader discussions of human rights and public benefit.[49]

Ethical issues are also inextricably linked with copyright concerns. The rapid development of AI has led many individuals to use such tools to create authorial texts and visual materials. A number of AI-created texts have already become bestsellers. Within academic contexts, both students and teachers increasingly turn to AI for help in writing qualification papers and scientific publications. At first glance, AI assistance, when used wisely, does not seem ethically problematic, since, for example, the use of algorithmic skills, information search, and calculations is quite reasonable and adequate. At the same time, many people use it to minimise their own participation in solving intellectual problems, which, in fact, is an ethical problem.

Moreover, AI-developed materials often lack a specific author but are actively sold on freelance platforms, which cannot currently refer to specific laws for detailed regulation of such cases. Traditional legal concepts of copyright do not apply to examples of artificial intelligence, since no specific legislative sanctions have been developed for it. As a result, it is unclear who should be punished in such a case. Lucci argues that traditional copyright concepts are outdated and require not just revision, but the creation of entirely new ones in the era of digitalisation and AI.[50]

A rather curious approach is proposed by Morley and her colleagues, who argue that it is necessary to develop legal norms that allow the creation of a copyright object without direct human participation. However, such a model raises additional questions regarding liability if an AI-generated work causes damage to someone. For example, a generated scene depicting violence may harm minors or extremely impressionable individuals.[51] Although such cases are currently rare—and AI platforms typically do not allow the creation of prompts that contain violence, or generate safe text/images—precedents may emerge.

If such a precedent arises, it seems logical to us to impose liability primarily on the platform owner where this situation occurred. This position is justified: when a platform acquires or deploys an AI mode, it enters into an agreement that sets out the rules for regulating relations with users. Accordingly, the AI developer should not be held liable, as responsibility is shifted to those who directly use the AI. In this case, it is necessary to develop a legal basis for regulating the conclusion of contracts when purchasing the right to place an AI model.

Researchers Moroz and Muzaparov proposed an interesting point on the ethical regulation of AI use.[52] According to these authors, the law on AI should take into account the specifics

48 de Almeida, dos Santos and Farias (n 27).
49 Idrysheva (n 19).
50 Lucchi (n 41).
51 Morley and others (n 47).
52 Moroz and Muzaparov (n 12).

of a specific AI platform, since many of the results of generations cannot be attributed to the concepts of "creativity" and "authorship". In view of this, it is necessary to develop and introduce new terminology in the context of AI to allow for a more understandable interpretation of the results of user interaction with artificial intelligence and, therefore, to regulate potential problems associated with authorship.

At this time, it is not possible to determine whether the Kazakh legislator will follow the path of harmonising national legislation with international standards or develop its own concept of ethical principles in the field of AI ethics. Still, international experience is nevertheless likely to facilitate the implementation of such technologies. Among the recommended postulates in the field of AI regulation, in general and its ethical foundations in particular, the norms and recommendations of the Global Digital Compact, planned for adoption at the UN level, would be significant for Kazakhstan and could provide the necessary guidelines for national regulation. However, even the implementation of international ethical standards in national AI legislation will be associated with a greater declaratory nature, and their voluntary compliance will often conflict with economic feasibility, since AI developers will often be guided by considerations of monetisation and applied usefulness first and foremost. However, it is hoped that national legislation will supplement AI's ethical postulates with normative regulation, balancing positive ethical guidelines with restrictive regulatory measures.

3.2. Institutional and Procedural Safeguards

The trend toward digitalising a significant portion of the public administration sector in the Republic of Kazakhstan is gaining momentum each year. The President of the country has articulated ambitious goals for transforming the state into a fully-fledged "digital nation," characterised by comprehensive digitalisation and the implementation of AI. In this context, there are stated intentions to introduce AI into various spheres of public administration, including geological exploration, monitoring of agricultural lands, transport and transit regulation, implementation of a multifunctional digital platform for transportation management, building information modeling using AI technologies, a digital platform for water resources, tax administration, the introduction of AI-based distance learning for rural regions, a system for monitoring the quality and volume of medical services using AI technologies, as well as the development of culture and the arts in the era of artificial intelligence, among others.[53]

The accelerated pace of digitalisation necessitates legal and regulatory measures in this field to ensure oversight, transparency in the functioning of digital systems, and the protection of citizens' rights, including procedural rights.

53 Tokayev (n 7).

As previously noted, the Law of the Republic of Kazakhstan "On Artificial Intelligence"[54] is characterised by excessive overregulation, the establishment of numerous administrative functions, and restrictions and prohibitions.

In particular, the Law "On Artificial Intelligence" establishes a classification of autonomous artificial intelligence systems (decision-making processes that are independent of predetermined parameters and are not subject to control by the system owner), as well as artificial intelligence systems with a specific set of functional capabilities (paragraph 2, Article 17). These include, in particular, the use of subconscious, manipulative, or other methods that significantly distort human behaviour; exploitation of human moral and/or physical vulnerability; determination of human emotions without consent; and similar practices.

These grounds largely mirror the provisions of the EU AI Act (Article 5),[55] albeit with certain distinctions.

In accordance with subparagraph 3 of paragraph 3 of Article 17 of the Law "On Artificial Intelligence", the use of AI systems that evaluate and classify natural persons or groups of persons over a certain period of time based on their social behavior or known, presumed, or predicted personal characteristics is prohibited, except in cases provided for by the laws of the Republic of Kazakhstan.

By contrast, under the EU AI Act, "social scoring" is prohibited only where it results in discrimination or disproportionate sanctions. The EU regulation sets out unambiguous conditions under which evaluations or classifications may be carried out. Accordingly, any normative acts of competent authorities aimed at creating an online platform/application using AI would be required to comply with the principle established in Article 5 of the EU AI Act.

The corresponding provision in the domestic legislation establishes an absolute prohibition, thereby creating a barrier to digital development. The introduction of any AI-based technologies designed to evaluate or classify individuals or groups (e.g., credit or social scoring) based on their social behaviour would require the adoption of a law, either through amendments to existing legislation or the enactment of a new law. This would create significant administrative hurdles, as the legislative process entails a specific sequence of steps, requirements, and timelines.

At the same time, the formulation "except in cases provided for by the laws of the Republic of Kazakhstan" may enable state bodies to be granted very broad powers. For instance, legislation may confer competence upon an authorised body to adopt corresponding subordinate normative legal acts. In such a case, there is a significant likelihood that the rights and legitimate interests of citizens could be infringed by the state's imperative

54 Law of the Republic of Kazakhstan No 230-VIII (n 9).

55 Regulation (EU) 2024/1689 (n 10).

administrative will. In this regard, it appears necessary to establish, at the legislative level, a condition requiring that any evaluation or classification of individuals or groups not lead to adverse or discriminatory treatment of particular persons or groups.

In addition, it seems necessary to introduce a requirement to protect citizens' fundamental rights. Any ratings or classifications based on social indicators must not restrict fundamental rights and freedoms, which include, *inter alia*, those enshrined in the Constitution of the Republic of Kazakhstan, including the right to life, personal liberty, the inviolability of private life, the right to recognition of legal personality, and judicial protection.

A different approach is applied to the classification of natural persons based on their biometric data. Under subparagraph 5 of paragraph 3 of Article 18, classifications aimed at drawing conclusions about race, political views, religious affiliation, or other circumstances are prohibited in the Republic of Kazakhstan if they are to be used for discriminatory purposes. By comparison, the EU AI Act permits the placing on the market, putting into service, or use of biometric categorisation systems that determine race, political opinions, trade union membership, religious or philosophical beliefs, sexual life or orientation, in cases of lawful labelling or filtering of biometric data, for instance, in the field of law enforcement.

In this regard, the formulation "and on any other grounds" makes it possible to encompass all potential bases of discrimination, including those not explicitly specified in current legislation. In the context of rapid technological development and the emergence of new forms of discriminatory practices, such openness ensures regulatory flexibility. The focus on the ultimate effect—"for the purpose of any discriminatory use"—allows the state not to block the very process of developing and testing technologies, but rather to concentrate on preventing unlawful consequences. This reduces the risk of excessive interference in scientific research or the neutral use of biometric data, for example, in medical or educational contexts.

Thus, the regulatory challenges under the Law "On Artificial Intelligence" manifest in excessive regulatory density, the creation of administrative barriers to technological deployment, and the risks of broad discretion by state authorities—factors that may slow digital development and threaten citizens' procedural rights.

3.3. Liability and Protection of Citizens' Rights

Continuing the discussion on the use of biometric data, it is impossible not to address the relationship between state regulation and the safeguarding of citizens' rights, including the establishment of appropriate liability for violations of their protection. The current Law of the Republic of Kazakhstan "On Personal Data and Its Protection" (hereinafter: the Law on Personal Data) classifies biometric data as personal data that characterises the

physiological and biological features of the subject of personal data, based on which his or her identity can be established. [56]

The protection of such data is, first and foremost, established through the recognition of its confidentiality. Thus, pursuant to Article 11 of the Law on Personal Data, owners and/or operators, as well as third parties obtaining access to personal data of restricted access, are required to ensure their confidentiality by adhering to the obligation not to disclose such data without the consent of the data subject or his/her legal representative, or in the absence of another lawful basis. Persons who have become aware of personal data of restricted access in connection with professional or official necessity, as well as through employment relations, are likewise obliged to ensure their confidentiality.

At the same time, the high-profile case involving Kaspi Bank regarding the breach of the confidentiality of digital data reveals gaps in the legislation, which, in turn, lead to violations of citizens' rights.[57]

In 2021, a citizen of Kazakhstan reported a privacy violation while applying for a loan through the Kaspi Bank mobile application, which required biometric identification. According to publicly available media sources, the user alleged that her personal data were processed and subsequently shared among debt collection agencies without her consent. Over several years, she appealed to the bank and relevant state authorities, requesting the deletion of the biometric data and clarification of responsibility. The bank, in its official response, stated that the data breach occurred on the side of third-party collectors and was not caused by its employees. It should be noted that no official court proceedings or other procedural documents are available for this case; however, certain elements of this precedent may still be analysed as an illustrative example.

Under paragraph 3 of Article 11 of the Law on Personal Data, the confidentiality of biometric data is established by the legislation of the Republic of Kazakhstan. The procedure for biometric identification carried out by banks is determined by the relevant Rules (hereinafter: the Rules).[58] These Rules clearly provide that biometric identification is to be conducted using the person's face (paragraph 5). Accordingly, obtaining images of any other body, apart from the individual's face, is unlawful and such data must neither be stored in the bank's database nor transferred to collection agencies.

56 Law of the Republic of Kazakhstan No 94-V 'On Personal Data and their Protection' (21 May 2013) <https://adilet.zan.kz/eng/docs/Z1300000094> accessed 12 September 2025.

57 Arman Ermekov, 'A Woman Accuses the Bank of Distributing her Intimate Photo' (*Politico*, 15 August 2025) <https://politico.kz/article/ayel-adam-kaspi-bankti-ashyk-suretin-taratqany-ushin-sotka-bermek> accessed 12 September 2025.

58 Resolution of the Board of the Agency of the Republic of Kazakhstan for Regulation and Development of the Financial Market No 56 'On approval of the Rules for Conducting Biometric Identification by banks, Organizations Carrying out Certain Types of Banking Operations, and Microfinance Organizations' (16 August 2024) <https://adilet.zan.kz/kaz/docs/V2400034950> accessed 12 September 2025.

At the same time, the Rules contain an important limitation. Pursuant to paragraph 1, they do not apply to biometric identification processes carried out by banks using their own hardware devices. In other words, the Rules are inapplicable when banks use their own devices—such as ATMs, terminals, in-branch systems, or mobile applications—as was the case in the aforementioned incident.

Pursuant to paragraph 5-5 of Article 34 of the Law of the Republic of Kazakhstan "On Banks and Banking Activities", banks are prohibited from concluding a bank loan agreement with an individual via the Internet without conducting biometric identification, the procedure for which must be determined by the authorised body.[59]

According to Kazakhstan's legislation, subordinate regulatory legal acts—such as rules, regulations, instructions, methodologies—do not establish norms of law but are adopted only to implement legislative acts and other higher-level normative legal acts. In the case of the Rules, it is evident that the provision granting banks the competence, in accordance with their internal regulations, to independently determine the procedure and requirements for biometric identification and the subsequent handling of digital data constitutes a norm of law and should be regulated at the statutory level. Otherwise, as illustrated by the above-mentioned case, banks may rely on internal regulations, commercial secrecy, or other legally protected information to violate citizens' rights and freedoms while evading responsibility for breaches of personal data legislation.

At present, the draft Digital Code of the Republic of Kazakhstan (hereinafter: the Code) is under consideration by the Mazhilis of the Parliament of the Republic of Kazakhstan.[60]

According to the Code, the following biometric data of citizens of the Republic of Kazakhstan will be subject to processing for authentication purposes:

1) digital facial image;

2) dactyloscopic (fingerprint) information.

Thus, the adoption of the Code is expected to eliminate the existing legal gap. Nevertheless, it remains necessary to review current legislation to establish more effective mechanisms for protecting individuals' personal data and safeguarding related civil rights and freedoms.

59 Law of the Republic of Kazakhstan No 2444 'On Banks and Banking Activities in the Republic of Kazakhstan' (31 August 1995) <https://adilet.zan.kz/kaz/docs/Z950002444_> accessed 12 September 2025.

60 Draft Digital Code of the Republic of Kazakhstan (2024) <https://mazhilis.parlam.kz/kk/all-bill/807> accessed 9 September 2025.

4 CONCLUSIONS

The list of problems outlined above will not be complete now or in the future, which reflects the evolving nature and inherent complexity of artificial intelligence. Given the current absence of AI-specific regulation in Kazakhstan, the country will need to develop its regulatory toolkit both through foreign experience and local regulation, taking into account domestic achievements and failures in the context of the underdevelopment of AI technologies and the lack of basic experience in its use.

An analysis of foreign experience enables the identification of three models of legal regulation of AI in Kazakhstan. The first model involves copying international experience by transferring advanced international norms to Kazakhstani legislation. The second model can be based on the symbiosis of international legal norms and national characteristics of the creation and use of AI technologies. The third approach is based on minimal legal regulation to obtain a technological advantage through soft regulation.

At the same time, comprehensive regulation of AI by norms of direct action, combining normative, technical, and ethical regulation, is difficult to implement and, most likely, even harmful to national legislation due to the multidimensionality of the phenomenon and the insufficiency of research and empirical data on artificial intelligence.

It appears likely that, upon adopting the Law "On Artificial Intelligence", the Kazakh legislator will also enact a number of related regulatory legal acts. These may address issues such as liability for violations of citizens' rights arising from the use of AI technologies, procedures for compensating for damage caused by AI or robots based on such technologies, copyright protection when creating products generated by AI technologies, and the ethical foundations of AI technologies. In this context, legal regulation will shift in favour of AI's well-known characteristics and the risks associated with it. Given the scope of regulation required, Kazakhstan's lawmakers cannot realistically anticipate all potential AI-related risks.

Any approach to forming an AI regulatory framework in Kazakhstan must, on the one hand, seek to balance the practical utility of AI technologies with the potential risks associated with their deployment and use, and on the other hand, select appropriate regulatory tools not only based on our own experience, but also a preventive analysis of existing technologies in other countries.

Thus, an analysis of the current situation regarding AI regulation in Kazakhstan yields several general conclusions. The creation of a legal framework for regulating AI should be accompanied by flexible legislative approaches when adopting foreign experience, especially from countries that lack a strong regulatory framework in the field of artificial intelligence but are actively implementing such technologies. The development of national standards for AI from an ideal perspective should be carried out by state research institutions, but in close cooperation with private developers of AI technologies, with additional verification with

developers of similar technologies abroad. The axiological approach to the development of ethical principles for the functioning of AI should be dominant, but intuitively accepted at the level of technology developers with the ability to revise the list of principles themselves without breaking away from the developing practice.

In the context of the differentiation of national legislations regulating artificial intelligence, the globalisation of uniform approaches to AI regulation is the most productive instrument of unification. At the same time, in the foreseeable future, full harmonisation of international norms with national ones across all countries is practically impossible, since the technological component of the economic development of interested countries is competitive and does not support conscious limitation of gains.

REFERENCES

1. Copeland BJ, 'History of Artificial Intelligence (AI): Alan Turing and the Beginning of AI', *Britannica* (7 November 2025) <https://www.britannica.com/science/history-of-artificial-intelligence> accessed 10 November 2025
2. de Almeida PGR, dos Santos CD and Farias JS, 'Artificial Intelligence Regulation: A Framework for Governance' (2021) 23 Ethics and Information Technology 505. doi:10.1007/s10676-021-09593-z
3. Floridi L, 'Introduction to the Special Issues: The Ethics of Artificial Intelligence: Exacerbated Problems, Renewed Problems, Unprecedented Problems' (2024) 61(4) American Philosophical Quarterly 301. doi:10.5406/21521123.61.4.01
4. Golm A, 'Tokayev Spoke about the Development of AI and the Creation of a Supercomputer in Kazakhstan' (*NUR.KZ*, 12 April 2024) <https://www.nur.kz/technologies/software/2083760-tokaev-vyskazalsya-o-razvitii-ii-i-sozdanii-superkompyutera-v-kazahstane/> accessed 20 August 2025
5. Hagendorff T, 'The Ethics of AI Ethics: An Evaluation of Guidelines' (2020) 30 Minds & Machines 99. doi:10.1007/s11023-020-09517-8
6. Hamada M and others, 'Artificial Intelligence to Improve the Business Efficiency and Effectiveness for Enterprises in Kazakhstan' (2021) 4(1) SAR Journal - Science and Research 34. doi:10.18421/SAR41-06
7. Idrysheva SK, 'On the Digital Code of Kazakhstan' (2022) 3(96) Law and State 72. doi:10.51634/2307-5201_2022_3_72
8. Kuandykova ES, Baideldinov DL and Hoffmann T, 'Problems of Legal Regulation of Digital Transformation of Agriculture of the Republic of Kazakhstan' (2023) 112(4) Bulletin of the Karaganda University: Law Series 7. doi:10.31489/2023L4/7-17
9. Lucchi N, 'ChatGPT: A Case Study on Copyright Challenges for Generative AI Systems' [2023] European Journal of Risk Regulation 1. doi:10.1017/err.2023.59

10. Morley J and others, 'Operationalising AI Ethics: Barriers, Enablers and Next Steps' (2023) 38 AI & Society 411. doi:10.1007/s00146-021-01308-8

11. Moroz S and Muzaparov S, 'Problems of Copyright and Intellectual Property Rights in Connection with the Use of AI Technologies (Neural Networks) (2024) 198 Scientific Collection InterConf 260

12. Murdoch B, 'Privacy and Artificial Intelligence: Challenges for Protecting Health Information in a New Era' (2021) 22 BMC Med Ethics 122. doi:10.1186/s12910-021-00687-3

13. Reier Forradellas RF and Garay Gallastegui LM, "Digital Transformation and Artificial Intelligence Applied to Business: Legal Regulations, Economic Impact and Perspective' (2021) 10(3) Laws 70. doi:10.3390/laws10030070

14. Siau K and Wang W, 'Artificial Intelligence (AI) Ethics: Ethics of AI and Ethical AI' (2020) 31(2) Journal of Database Management 74. doi:10.4018/JDM.2020040105

15. Suleimenov MK and Karagusov FS, "The Concept of Recodification of the Civil Code of Ukraine and Modernization of the Civil Code of Kazakhstan: A Comparative Analysis of the Main Ideas" (*Paragraph Lawyer*, 29 July 2021) <https://online.zakon.kz/Document/?doc_id=32892885> accessed 20 August 2025.

16. Tlembayeva ZU, 'On Legal Regulation of the Use of Artificial Intelligence in Healthcare of the Republic of Kazakhstan' (2022) 5-1 Greater Eurasia: Development, Security, Cooperation 1123.

17. Tlembayeva ZU, 'On Some Approaches to the Legal Regulation of Artificial Intelligence' (2021) 2(65) Bulletin of the Institute of Legislation and Legal Information of the Republic of Kazakhstan 61. doi:10.52026/2788-5291_2021_65_2_61

18. Tokayev KJ, 'Kazakhstan in the Era of Artificial Intelligence: Current Challenges and Solutions through Digital Transformation: President's State of the Nation Address to the People of Kazakhstan' (*Әділет*, 8 September 2025) <https://adilet.zan.kz/kaz/docs/K25002025_1> accessed 10 November 2025

19. Zhanysbayeva D and Ryzhkina M, 'Draft Law of the Republic of Kazakhstan "On Artificial Intelligence" – Principles of Regulation and Practical Aspects' (*GRATA International*, 14 July 2025) <https://gratanet.com/publications/draft-law-of-the-republic-of-kazakhstan-on-artificial-intelligence-principles-of-regulation-and-practical-aspects> accessed 9 September 2025

AUTHORS INFORMATION

Anuar Nurmagambetov*
Candidate of Sciences (Law), Assoc. Prof., Kokshetau University named after Abay Myrzakhmetov, Kokshetau, Kazakhstan
an.nurmagambetov@gmail.com
https://orcid.org/0009-0006-4362-9556
Corresponding author, responsible for conceptualization, data curation, methodology, resources, validation, writing – original draft, writing – review & editing.

Anet Nurmagambetov
PhD (Law), Assoc. Prof., L.N. Gumilyov Eurasian National University, Astana, Kazakhstan
anetnurmagambetov@gmail.com
https://orcid.org/0000-0002-8909-4618
Co-author, responsible for data curation, methodology, resources, validation, writing – original draft, writing – review & editing.

Amanzhol Nurmagambetov
Dr.Sc. (Law), Prof., Higher Law School, Astana International University, Astana, Kazakhstan
amanzholnurmagambetov@gmail.com
https://orcid.org/0000-0001-9026-9019
Co-author, responsible for data curation, methodology, resources, supervision, validation, writing – original draft, writing – review & editing.

Aigerim Zhumabayeva
PhD (Law), Department of Material and Technical Support, REM «Institute of Parliamentarism», Astana, Kazakhstan
aigerimzhumabayeva9933@gmail.com
https://orcid.org/0000-0003-3376-4325
Co-author, responsible for data curation, methodology, resources, validation, writing – original draft, writing – review & editing.

Competing interests: No competing interests were disclosed. Any potential conflict of interest must be disclosed by authors.

Disclaimer: The authors declare that their opinion and views expressed in this manuscript are free of any impact of any organizations.

RIGHTS AND PERMISSIONS

EDITORS

ABOUT THIS ARTICLE

Cite this article
Nurmagambetov Anuar, Nurmagambetov Anet, Nurmagambetov Amanzhol and Zhumabayeva A, 'Artificial Intelligence and Law: Procedural Safeguards and Regulatory Challenges in Kazakhstan' (2025) 8(Spec) Access to Justice in Eastern Europe 119-45 <https://doi.org/10.33327/AJEE-18-8.S-a000157>

DOI: https://doi.org/10.33327/AJEE-18-8.S-a000157

Keywords: *artificial intelligence, legal regulation of artificial intelligence, object and subject of artificial intelligence, international regulation in the field of artificial intelligence, artificial intelligence in Kazakhstan.*

DETAILS FOR PUBLICATION

Date of submission: 25 Aug 2025
Date of acceptance: 28 Oct 2025
Online First publication: 04 Dec 2025
Last Published: 30 Dec 2025
Whether the manuscript was fast tracked? - No
Number of reviewer report submitted in first round: 3 reports (2 external reviewers and 1 guest editor)
Number of revision rounds: 2 rounds with minor revisions

Technical tools were used in the editorial process:
Plagiarism checks - Turnitin from iThenticate https://www.turnitin.com/products/ithenticate/
Scholastica for Peer Review https://scholasticahq.com/law-reviews

AI DISCLOSURE STATEMENT

The corresponding author confirmed that AI technologies have only been used to enhance language clarity and grammar. No AI tools were used to generate ideas, structure arguments, analyze data, or produce conclusions.

АНОТАЦІЯ УКРАЇНСЬКОЮ МОВОЮ

Дослідницька стаття

ШТУЧНИЙ ІНТЕЛЕКТ І ПРАВО: ПРОЦЕСУАЛЬНІ ГАРАНТІЇ ТА РЕГУЛЯТОРНІ ВИКЛИКИ В КАЗАХСТАНІ

Ануар Нурмаґамбетов*, Анет Нурмаґамбетов, Аманжол Нурмаґамбетов та Айґерім Жумабаєва

АНОТАЦІЯ

Вступ. *Активна інтеграція штучного інтелекту (ШІ) у різноманітні сфери людської діяльності створила значні можливості для інновацій та підвищення ефективності, водночас породжуючи складні етичні, правові та соціальні проблеми. Серед них особливої уваги потребує впровадження систем ШІ з високим рівнем ризику через їхній потенційний вплив на основоположні права людини, громадську безпеку та соціально-економічні відносини. У цьому дослідженні розглядаються як переваги, так і ризики технологій ШІ з наголосом на необхідності створити чітку правову та нормативну базу на національному та міжнародному рівнях.*

Методи. *У дослідженні використовується порівняльно-правовий аналіз наявних регуляторних підходів, зокрема Акт Європейського Союзу про штучний інтелект (Акт ЄС про ШІ), Принципи ШІ щодо ОЕСР та національну законодавчу практику. Методологія ґрунтується на систематичному перегляді нормативно-правових актів, доктринальних джерел і аналітичних документів, а також на оцінці потенційних ризиків, пов'язаних з використанням систем ШІ з високим рівнем ризику в різних сферах, зокрема, що стосується транспорту, охорони здоров'я та фінансових послуг.*

Результати та висновки. *Аналіз показує, що хоча впровадження штучного інтелекту сприяє економічному розвитку, ефективності державного управління та покращує якість послуг, воно також породжує такі ризики, як дискримінація, порушення конфіденційності, кіберзагрози та зниження рівня відповідальності. Зокрема, у дослідженні підкреслюється, що чинне законодавство в Казахстані, як і в багатьох інших юрисдикціях, недостатньо враховує особливості систем штучного інтелекту з високим рівнем ризику. Водночас порівняльно-правовий аналіз показує, що найбільш ефективні моделі регулювання*

ґрунтуються на ризик-орієнтованому підході, забезпечуючи прозорість, людський контроль та механізми відповідальності. Результати дослідження свідчать про те, що часткові зміни до чинного законодавства (наприклад, у сфері обов'язкового страхування та захисту прав споживачів) можуть слугувати тимчасовим заходом, тоді як ухвалення спеціального закону про ШІ є неминучим з огляду на довгострокову перспективу.

Дослідження підкреслює необхідність створення збалансованої правової бази, яка гармонізує технологічні інновації із захистом прав людини та суспільних інтересів. Стверджується, що Казахстан, враховуючи кращий міжнародний досвід, повинен дотримуватися двоетапного підходу: (1) внесення цільових змін до галузевого законодавства; і (2) розробка комплексного закону про ШІ, зосередженого на системах з високим рівнем ризику. Така структура зменшить ризики, забезпечить підзвітність і сприятиме суспільній довірі, одночасно заохочуючи до відповідального та сталого використання штучного інтелекту.

Ключові слова: *штучний інтелект, правове регулювання штучного інтелекту, об'єкт і суб'єкт штучного інтелекту, міжнародне регулювання у сфері штучного інтелекту, штучний інтелект у Казахстані.*

Access to Justice in Eastern Europe
ISSN 2663-0575 (Print)
ISSN 2663-0583 (Online)
http://ajee-journal.com

Research Article

DIGITAL RIGHTS, AI, AND THE LAW: INTERNATIONAL PERSPECTIVES ON SAUDI ARABIA'S LEGAL FRAMEWORK AND INTERNATIONAL EXPERIENCE

Soumaya Khammassi and Yusra AlShanqityi*

ABSTRACT

Background: *This research analyses the global evolution of digital rights and AI governance and examines the implications for Saudi Arabia's legal framework. As artificial intelligence becomes increasingly embedded in everyday life, there is an urgent need to assess its impact on fundamental human rights, particularly given the absence of a global consensus on the definition and scope of "digital human rights." The Kingdom of Saudi Arabia's (KSA) Vision 2030 initiative presents a unique opportunity to develop AI governance frameworks that align with international standards while reflecting local cultural values and Islamic ethical principles.*

Methods: *This study employs a qualitative analytical approach to examine Saudi Arabia's current legal framework governing AI and digital human rights. The methodology involves a comprehensive statutory analysis of Saudi Arabian legislation, particularly the Personal Data Protection Law (PDPL) and the Saudi Data and AI Authority (SDAIA) 's ethical guidelines, in comparison to international legal instruments, including*

DOI:
https://doi.org/10.33327/AJEE-18-8.S-a000154

Date of submission: 16 Apr 2025
Date of acceptance: 08 Sep 2025
Online First publication: 08 Dec 2025
Last Published: 30 Dec 2025

Disclaimer:

The authors declare that their opinion and views expressed in this manuscript are free of any impact of any organizations.

UNESCO's AI Ethics Recommendation, the EU AI Act, and OECD guidelines. The research evaluates alignment with international standards and effectiveness in addressing emerging digital rights challenges in the AI era.

Results and conclusions: *The research reveals that while Saudi Arabia has made notable progress through the PDPL and SDAIA frameworks, significant regulatory gaps persist. The PDPL exhibits limitations in addressing contemporary AI challenges, including algorithmic accountability, bias mitigation, and comprehensive data protection in AI contexts. The study identifies critical deficiencies, including overly broad exceptions to consent requirements, insufficient provisions for algorithmic transparency, and fragmented regulatory oversight. Recommendations include establishing specialised oversight bodies, developing ethical frameworks tailored to the Saudi context, and increasing the involvement of experts in decisions related to AI governance. This paper contributes by offering a comparative evaluation of Saudi Arabia's digital rights framework against leading international instruments, highlighting reforms necessary for culturally grounded and globally aligned governance.*

1 INTRODUCTION

1.1. Research Context and Significance

Artificial intelligence is a trending issue transforming multiple sectors, including healthcare, education, governance, and societal systems. Researchers increasingly recognise both the positive and negative impacts AI can have on human rights, particularly digital human rights.[1] The implementation of AI raises significant ethical and legal issues regarding privacy, fairness, and transparency. As AI technologies advance, specific legislation becomes necessary to ensure they respect individual rights in the digital world.

At the core of the current technological evolution centred on AI's transformative potential, growing concerns exist about its effects on human rights. The legal debate primarily addresses three fundamental issues: how these new technologies violate existing rights, create conflicts among established rights, and introduce entirely new legal questions for which no established rights framework yet exists. This complex intersection of technology and rights requires thoughtful consideration of equity, transparency, and accountability in AI governance frameworks.[2]

The Kingdom faces a distinctive opportunity to contribute to the global discourse on digital rights by articulating an approach that harmonises international norms with

1 Reema Bakheet Alzahrani, 'An Overview of AI Data Protection in the Context of Saudi Arabia' (2024) 3(3) International Journal for Scientific Research 199. doi:10.59992/IJSR.2024.v3n3p8 [in Arabic].

2 Bart Custers, 'New Digital Rights: Imagining Additional Fundamental Rights for the Digital Era' (2022) 44 *Computer Law & Security Review* 105636. doi:10.1016/j.clsr.2021.105636.

Islamic ethical principles and local cultural values. Saudi Arabia's Vision 2030[3] initiative explicitly recognises the transformative potential of digital technologies, but this transformation necessitates careful consideration of how rights frameworks must evolve in parallel. Vision 2030 aims to diversify the economy, rebalance the country's dependence on oil revenues, and develop new technologies. This has led to setting strategic objectives for the digitalisation of several sectors, including the application of artificial intelligence and data technologies in health and other sectors.[4] However, as machine intelligence is increasingly implemented across governance activities and multiple aspects of life, there is a need to address human rights in the digital world. As a modernising country in the contemporary world, the KSA stands at the intersection of technology and human rights in the digital environment.

Unlike countries where digital rights discussions emerged gradually alongside technological development, K.S.A is engaging with these questions during an accelerated period of digital transformation, potentially allowing for more integrated approaches to rights protection within technological infrastructure.[5]

1.2. Research Problem and Gap

The governance of artificial intelligence at the international level has evolved rapidly, moving from early ethical principles to increasingly formalised regulatory approaches. The Asilomar AI Principles (2017)[6] and IEEE's "Ethically Aligned Design" initiative (2016-2019)[7] represented initial efforts from the research and engineering communities.[8] More comprehensive intergovernmental frameworks emerged with the OECD Principles on AI (2019)[9] and UNESCO's Recommendation on the Ethics of AI (2021),[10] adopted by 193 countries.

3 *Saudi Vision 2030* (2025) <https://www.vision2030.gov.sa/en> accessed 10 April 2025.

4 Custers (n 2).

5 Mohammad Omar Mohammad Alhejaili, 'Integrating Smart Contracts into the Legal Framework of Saudi Arabia' (2025) 67(2) International Journal of Law and Management 230. doi:10.1108/IJLMA-03-2024-0086.

6 Future of Life Institute, 'Asilomar AI Principles' (2017) <https://futureoflife.org/ai-principles/> accessed 10 April 2025.

7 IEEE, *Ethically Aligned Design: A Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems* (IEEE 2019); IEEE, 'The IEEE Global Initiative 2.0 on Ethics of Autonomous and Intelligent Systems' (*IEEE Standards Association (IEEE SA)*, 2025) <https://standards.ieee.org/content/ieee-standards/en/industry-connections/ec/autonomous-systems.html> accessed 10 April 2025.

8 Alan FT Winfield and Marina Jirotka, 'Ethical Governance is Essential to Building Trust in Robotics and Artificial Intelligence Systems' (2018) 376(2133) Philosophical Transactions of the Royal Society A 20180085. doi:10.1098/rsta.2018.0085.

9 OECD, 'Recommendation of the Council on Artificial Intelligence' (*OECD Legal Instruments*, 22 May 2019) <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449> accessed 10 April 2025.

10 UNESCO, 'Recommendation on the Ethics of Artificial Intelligence' (*UNESCO Digital Library*, 2021) <https://unesdoc.unesco.org/ark:/48223/pf0000380455 > accessed 10 April 2025.

International frameworks exist along a spectrum from voluntary principles to legally enforceable regulations. The OECD Principles and UNESCO Recommendation function as "soft law," shaping behaviour through normative expectations rather than formal sanctions.[11] The EU AI Act[12] establishes legally binding obligations with significant penalties for non-compliance.[13] However, no universally accepted definition of digital human rights exists, and no binding treaty framework comprehensively addresses them.

Within this context, K.S.A presents a particularly compelling case: the Kingdom is undergoing accelerated digital transformation under Vision 2030 while integrating Islamic ethical principles into its legal system. Research examining how K.S.A's domestic framework aligns with, diverges from, or innovates upon international digital rights governance remains limited.

1.3. Research Objectives and Questions

This paper addresses the previously explained gap by critically examining K.S.A's evolving legal framework on digital rights and AI governance in light of international standards. It is guided by the following research questions:

1. To what extent are digital human rights conceptually established within international legal discourse and scholarly literature?
2. How has Saudi Arabia integrated global digital human rights frameworks and principles into its domestic legislative and regulatory infrastructure?
3. What distinctive challenges and strategic opportunities characterise Saudi Arabia's approach to digital rights protection within the context of accelerating technological advancement?

11 Joel R Reidenberg, 'Lex Informatica: The Formulation of Information Policy Rules Through Technology' (1997) 76(3) Texas Law Review 553.

12 Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) [2024] OJ L 1689 <http://data.europa.eu/eli/reg/2024/1689/oi> accessed 10 April 2025.

13 UK Department for Digital, Culture, Media and Sport, 'Establishing a Pro-Innovation Approach to Regulating AI: An Overview of the UK's Emerging Approach' (*UK Government*, 18 July 2022) <https://www.gov.uk/government/publications/establishing-a-pro-innovation-approach-to-regulating-ai> accessed 10 April 2025.

1.4. Contribution and Structure

By answering these questions, this paper contributes to the growing body of literature on digital rights. It employs a comparative doctrinal analysis, focusing on statutory analysis, soft law examination, and jurisdictional analysis of the digital rights framework.[14]

The comparative doctrinal approach allows for positioning Saudi Arabia's regulatory efforts within a broader international framework.

This paper is organised into five main sections. Section 2 develops the conceptual foundation by examining how digital human rights are defined in international scholarship and identifying five core categories: privacy, internet access, data protection, anonymity, and the right to be forgotten. Section 3 situates these concepts within the MENA regional context, examining how Islamic legal and ethical principles interact with emerging digital governance frameworks. Section 4 analyses international AI governance approaches, contrasting the EU's binding regulatory model with UNESCO and OECD soft-law frameworks to establish comparative benchmarks. Section 5 evaluates Saudi Arabia's legal framework—primarily the PDPL and Anti-Cybercrime Law—against these international standards, identifying both achievements and regulatory gaps. The paper concludes with targeted recommendations for legal reform that balance international alignment with cultural distinctiveness.

It highlights Saudi Arabia's progress in AI regulation through a systematic analysis of existing legal instruments, such as the Personal Data Protection Law (PDPL) and the Saudi Data and Artificial Intelligence Authority (SDAIA) Ethical Guidelines. It helps to analyse their compliance with international standards and their ability to respond to new digital human rights issues in the era of AI. More broadly, the study underscores how cultural and religious values, particularly Islamic ethical principles, can inform distinctive approaches to digital rights governance in non-Western contexts.

2 LITERATURE REVIEW AND CONCEPTUALIZING DIGITAL HUMAN RIGHTS

2.1. Overview and Rationale

This section reviews the evolving concept of digital human rights, surveys the main categories identified in the literature, and examines international and regional governance

14 Royal Decree of the Kingdom of Saudi Arabia No M/19 of 09/02/1443 AH (16/09/2021 G) 'Personal Data Protection Law' (SDAIA 2023) <https://sdaia.gov.sa/en/SDAIA/about/Documents/Personal%20Data%20English%20V2-23April2023-%20Reviewed-.pdf> accessed 10 April 2025; Saudi Data and Artificial Intelligence Authority, *AI Ethics Principles* (SDAIA 2025) <https://sdaia.gov.sa/en/SDAIA/about/Documents/ai-principles.pdf> accessed 10 April 2025; Nayera Mohamed Hamed Ibrahim, 'Artificial Intelligence (AI) and Saudi Arabia's Governance' (2024) 40(4) Journal of Developing Societies 500. doi:10.1177/0169796X241288590.

approaches. The review highlights definitional debates, recurring principles, and regulatory models. It underscores the role of cultural and religious values, specifically Islamic ethical principles, in shaping non-Western approaches to digital rights governance.

2.2. Defining Digital Human Rights

The concept of digital human rights remains relatively new and emerging, unlike conventional human rights vested under International Human Rights frameworks such as the Universal Declaration of Human Rights (UDHR) and the International Covenant on Civil and Political Rights (ICCPR).[15] It was only during the early 2000s that international recognition of this category of rights began to emerge, with the World Summit on the Information Society (2003, 2005) marking a significant milestone in their institutional recognition.[16]

The digital human rights framework has evolved as an extension of traditional human rights frameworks, adapted to address the unique challenges posed by technological advancement. Digital rights are generally understood as human and legal rights that allow individuals to access, use, create, and publish digital content on devices such as computers and mobile phones, as well as in virtual spaces and communities.

The academic discussion reveals a significant question: do effective digital rights protections require entirely new frameworks, or can they be achieved by extending existing approaches? The answer to this issue remains challenging, since there is no agreement on a definition of digital human rights. While some scholars, such as Pangrazio and Sefton-Green,[17] define digital rights as merely a contextual expression of established rights, particularly freedom of expression and the right to privacy.[18] Others consider them a distinct category of rights,[19] particularly as new technologies like AI present issues not foreseen in traditional frameworks; thus, human rights are facing novel challenges that require new protection mechanisms.

15 Universal Declaration of Human Rights (UDHR) (10 December 1948) <https://www.un.org/en/about-us/universal-declaration-of-human-rights> accessed 10 April 2025; International Covenant on Civil and Political Rights (ICCPR) (16 December 1966) <https://treaties.un.org/pages/viewdetails.aspx?chapter=4&clang=_en&mtdsg_no=iv-4&src=ind> accessed 10 April 2025.

16 Rikke Frank Jørgensen, *Framing the Net: The Internet and Human Rights* (Edward Elgar 2013). doi:10.4337/9781782540809.

17 Luci Pangrazio and Julian Sefton-Green, 'Digital Rights, Digital Citizenship and Digital Literacy: What's the Difference?' (2021) 10(1) *Journal of New Approaches in Educational Research* 15. doi:10.7821/naer.2021.1.616.

18 Oleksandr M Kostenko and others, '"Legal Personality" of Artificial Intelligence: Methodological Problems of Scientific Reasoning by Ukrainian and EU Experts' (2023) 39(4) AI and Society 1683. doi:10.1007/s00146-023-01641-0.

19 Mireille Hildebrandt, *Smart Technologies and the End(s) of Law* (Edward Elgar 2015).

The definitional ambiguity surrounding digital human rights, emphasised by AllahRakha,[20] Bendary and Rajadurai,[21] undermines consistent protection because jurisdictions adopt divergent interpretations of these rights. To address this, scholars propose different strategies. Cong advocates a "translation approach," arguing that existing human rights principles remain valid but must be carefully reinterpreted for digital contexts.[22] Similarly, Land and Aronson highlight the potential of adapting traditional human rights frameworks to guide the governance of new technologies, reinforcing the idea that established principles can provide a normative foundation.[23] By contrast, Hildebrandt[24] insists that computational technologies introduce fundamentally novel challenges that cannot be resolved through reinterpretation alone. She cautions that smart technologies may erode the very "ends of law", justice, accountability, and legal certainty, unless new forms of "legal protection by design" are embedded directly into technological infrastructures.

Adding to this debate, Custers goes further by exploring the possibility of articulating entirely new fundamental rights for the digital era, arguing that incremental reinterpretation may not be sufficient to protect individuals against AI-driven risks.[25] More recent scholarship in the non-Western context expands this conversation by emphasising the role of cultural and religious traditions. For example, Elmahjub[26] proposes a pluralist ethical benchmarking for AI governance that incorporates Islamic ethics, while Ali et al.[27] explicitly evaluate AI through the lens of *maqāṣid al-sharīʿa*, demonstrating how Islamic jurisprudential principles of dignity, justice, and privacy can enrich digital rights discourse.[28] These perspectives underline that the recognition and definition of digital human rights should not only deal with whether they are "translated" or "new" rights, but also consider how cultural and religious frameworks may inform distinctive and contextually legitimate approaches to governance.

20 Naeem AllahRakha, 'UNESCO's AI Ethics Principles: Challenges and Opportunities' (2024) 2(9) *International Journal of Law and Policy* 24. doi:10.59022/ijlp.225.

21 Mohamed G Bendary and Jegatheesan Rajadurai, 'Emerging Technologies and Public Innovation in the Saudi Public Sector: An Analysis of Adoption and Challenges Amidst Vision 2030' (2024) 29(1) The Innovation Journal: The Public Sector Innovation Journal 1.

22 Wanshu Cong, 'Understanding Human Rights on the Internet: An Exercise of Translation?' (2017) 22(1-2) Tilburg Law Review 138. doi:10.1163/22112596-02201007.

23 Molly K Land and Jay D Aronson (eds), *New Technologies for Human Rights Law and Practice* (CUP 2018) doi:10.1017/9781316838952.

24 Hildebrandt (n 19).

25 Custers (n 2).

26 Ezieddin Elmahjub, 'Artificial Intelligence (AI) in Islamic Ethics: Towards Pluralist Ethical Benchmarking for AI' (2023) 36 Philosophy & Technology 73. doi:10.1007/s13347-023-00668-x.

27 Fatima Ali and others, 'Islamic Ethics and AI: An Evaluation of Existing Approaches to AI using Trusteeship Ethics' (2025) 38(2) Philosophy & Technology 120. doi:10.1007/s13347-025-00922-4.

28 Mohammad Omar Mohammad Alhejaili, 'Securing the Kingdom's e-Commerce Frontier: Evaluation of Saudi Arabia's Cybersecurity Legal Frameworks' (2024) 13(2) Journal of Governance & Regulation 275. doi:10.22495/jgrv13i2siart4.

Thus, digital rights are the rights individuals have in the digital realm. These rights are derived from conventional universal rights but are adapted to meet the demands and opportunities posed by the rapidly growing new technologies, including artificial intelligence. As digital environments increasingly shape humans' sociopolitical, economic, and personal existence, the scope of these rights has extended.

Ultimately, digital human rights represent a renewed paradigm, well-anchored in international human rights conventions, while suggesting an adaptive conceptual and normative approach.

2.3. Core Categories of Digital Human Rights

Comparative studies reveal both convergence and divergence across existing frameworks. Jobin et al., in their analysis of 84 ethics guidelines, identified documents transparency, justice, non-maleficence, responsibility, and privacy as recurring principles.[29] While the OECD emphasises inclusive growth and human-centred values, UNESCO places greater emphasis on cultural contexts and sustainability. The EU's AI Act adopts a more regulatory stance through its risk-based classification system.[30]

These frameworks address the issue of individuals' inherent rights in digital spaces with varying levels of specificity. Privacy protections are dominant in both ethical and legal discourse, though conceptualised differently across contexts. Non-discrimination principles appear consistently, but with varying approaches to bias mitigation. Transparency requirements have converged around key elements, including disclosure of AI use and appropriate explainability, though implementation guidance varies substantially.[31] Despite these variations, several consensus principles have emerged across frameworks. Fjeld et al. identified eight key themes with widespread support: privacy, accountability, safety and security, transparency and explainability, fairness and non-discrimination, human control of technology, professional responsibility, and promotion of human values.[32]

Taken together, these various frameworks show that digital rights discourse both continues established human rights traditions and introduces innovative protections. Building on this literature, the present research focuses on five essential digital rights: privacy, internet access, data protection, anonymity, and the right to be forgotten.

29 Anna Jobin, Marcello Ienca and Effy Vayena, 'The Global Landscape of AI Ethics Guidelines' (2019) 1(9) Nature Machine Intelligence 389. doi:10.1038/s42256-019-0088-2.

30 Michael Veale and Frederik Zuiderveen Borgesius, 'Demystifying the Draft EU Artificial Intelligence Act' (2021) 22(4) Computer Law Review International 97. doi:10.9785/cri-2021-220402.

31 Stefan Larsson and Fredrik Heintz, 'Transparency in Artificial Intelligence' (2020) 9(2) Internet Policy Review. doi:10.14763/2020.2.1469.

32 Gessfhk Fjeld and others, *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI* (Berkman Klein Center for Internet & Society Research Publication Series no 2020-1, Harvard University 2020). doi:10.2139/ssrn.3518482.

2.3.1. Privacy Rights in Digital Contexts

Digital privacy right is, ultimately, an extension of the privacy right. The international legal framework has long recognised privacy as a fundamental right. According to the International Covenant on Civil and Political Rights, Article 17 protects individuals against "arbitrary or unlawful interference with their privacy, family, home, or correspondence."[33] Article 12 of the Universal Declaration of Human Rights similarly asserts the right to privacy and protection from arbitrary interference.[34]

The problem does not lie in recognising the normative value of this right nor in defining its limits within the traditional framework in which it was established, but rather in its rapid transposition to the digital context. In the physical world, it is relatively easy to define the scope of application of this right, limit the actions considered violations, and assign ongoing responsibility for them. However, in the virtual realm, parameters become increasingly blurred. The decentralised nature of digital environments has shifted the focus from protecting individuals from surveillance to empowering them, raising concerns about digital access, inclusion, and protection and giving rise to profound regulatory uncertainty.

The use of advanced technologies, such as facial recognition, big data, and AI-assisted surveillance, has been unparalleled in human history and has been carried out primarily without people's knowledge or permission, thereby altering this right. For instance, in *Glukhin v. Russia* (App. No 11519/20),[35] the European Court of Human Rights stressed that the use of facial recognition to identify to identify, locate, and subsequently arrest the applicant from photographs and video posted on social media constituted an interference with his right to respect for his private life and infringe Article 8 of the European Convention on Human Rights.

Custers[36] expressively affirms that AI surveillance impacts privacy rights, leading to their violation, and calls for greater attention to ensuring that artificial intelligence adheres to privacy-freedom governing principles.[37] Furthermore, research on algorithmic bias is especially concerning for diverse societies. Studies from the United States demonstrate that even the most advanced AI facial recognition tools exhibit significantly higher error rates when identifying black women compared to white men.[38]

33 ICCPR (n 15) art 17; UN Human Rights Committee, 'CCPR General Comment No 16: Article 17 (Right to Privacy)' (8 April 1988) <https://www.refworld.org/legal/general/hrc/1988/en/27539> accessed 10 April 2025.

34 UDHR (n 15) art 12.

35 *Glukhin v Russia* App no 11519/20 (ECtHR, 4 July 2023) <https://hudoc.echr.coe.int/eng?i=001-225655> accessed 10 April 2025.

36 Custers (n 2).

37 Alzahrani (n 1).

38 Adam Schwartz, 'Chicago's Video Surveillance Cameras: A Pervasive and Poorly Regulated Threat to Our Privacy' (2013) 11(2) Northwestern Journal of Technology and Intellectual Property 47.

2.3.2. Right to Internet Access

The concept of a "right" to internet access, or "digital connectivity," implies the ability of a person to connect to the internet to seek, receive and impart information.[39] Although, there is no binding international treaty that recognises the right to internet, the internet is growingly considered as an enabler of many human rights such as the right to freedom of speech and expression and the right to information as provided under Article 19 of the Universal Declaration of Human Rights, or the right to have equal access to public services (Article 21).[40] Recognising that internet access plays a facilitative role in the enjoyment of fundamental rights, the UNHRC (United Nations Human Rights Council) adopted Resolution 47/16 in 2021, calling on governments to close the gap in the availability and accessibility of affordable, stable Internet.[41]

Scholarly perspectives on this issue can be divided into three distinct groups for this research. Advocates for classifying internet access as a human right emphasise how digital connectivity has become essential for exercising numerous established rights in contemporary society, from freedom of expression to education.[42] According to the second group, represented by scholars such as Veale and Zuiderveen Borgesius,[43] internet access is only a means to the enjoyment of other rights; consequently, they contend that it lacks the fundamental quality required for recognition as a human right. For this group, internet access serves merely as a technological enabler of rights rather than constituting a right itself.[44]

Among these perspectives, the most compelling is the view that internet access can be conceptualised as an emerging human right.[45] This nuanced approach acknowledges the internet's crucial role in contemporary rights fulfilment while acknowledging that it does not carry the same foundational status as primary human rights, such as freedom from torture or the right to life. Yet, in a world where digital infrastructure has become indispensable for daily life, meaningful participation in society becomes severely limited without internet access.

The resulting legal discourse reflects both continuity and innovation: it builds upon existing treaty obligations while pressing towards recognition of a distinct digital entitlement in practice.

39 Jonathon Penney, 'Open Connectivity, Open Data: Two Dimensions of the Freedom to Seek, Receive and Impart Information in the New Zealand Bill of Rights' (2012) 4 Victoria University of Wellington Law Review: Working Paper Series 1; Paul De Hert and Dariusz Kloza, 'Internet (access) as a New Fundamental Right: Inflating the Current Rights Framework?' (2012) 3(3) European Journal of Law and Technology 1.

40 UDHR (n 15) arts 19, 21.

41 Alhejaili (n 28).

42 Stephen Tully, 'A Human Right to Access the Internet? Problems and Prospects' (2014) 14(2) Human Rights Law Review 175. doi:10.1093/hrlr/ngu011.

43 Veale and Zuiderveen Borgesius (n 30).

44 Vinton Cerf, 'Internet Access is Not a Human Right' *The New York Times* (New York, 4 January 2012) A25.

45 Kay Mathiesen, 'The Human Right to Internet Access: A Philosophical Defense' (2012) 18 The International Review of Information Ethics 9. doi:10.29173/irie299.

2.3.3. Data Protection Rights

Data protection has become indispensable as AI systems process vast amounts of personal information, raising fundamental rights concerns. Traditionally, data protection has been governed by various legislations—such as the GDPR[46] in the EU—which grant individuals power and meaningful control over their information through rights of access, correction, and deletion.[47]

Data Protection right was firmly established in *Digital Rights Ireland Ltd v. Minister for Communications* (Joined Cases C-293/12 and C-594/12),[48] where the Court of Justice declared the Data Retention Directive invalid for exceeding the limits of proportionality under Articles 7, 8, and 52(1) of the Charter of Fundamental Rights. While acknowledging the Directive's legitimate objective of combating serious crime, the Court held that its general and indiscriminate retention of telecommunications data constituted a particularly serious interference with the rights to privacy and data protection. The judgment required any data retention regime to be strictly necessary, supported by clear and precise rules, and subject to independent supervision, thereby establishing the modern constitutional standard for data protection within the European Union.

However, the rise of AI places these standards under unprecedented pressure. The automated, predictive, and often opaque processing models that characterise AI challenge the very assumptions of informed consent, transparency, and proportionality that the Digital Rights Ireland judgment sought to safeguard. AI technologies rely on big data analytics and machine learning algorithms that often operate beyond conventional privacy safeguards. The "black box" nature of many AI systems—characterised by opaque decision-making processes lacking self-explainability—prevents individuals from understanding how their personal information is processed, analysed, and potentially shared.

This lack of transparency fundamentally weakens the principle of informed consent that underpins modern data protection regimes. As Vogel[49] argues, these obstacles can only be

46 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L 119/1.

47 Federica Paolucci, 'Enhancing Oversight and Addressing Gaps: Assessing the Impact of the AI Act on Biometric Identification Systems' in Natalia Menéndez González and Giuseppe Mobilio (eds), *Next Democratic Frontiers for Facial Recognition Technology (FRT): The Legal, Ethical and Democratic Implications of FRT* (Springer 2025) 71. doi.org/10.1007/978-3-031-89794-8_5.

48 Joined Cases C-293/12 and C-594/12 *Digital Rights Ireland Ltd v Minister for Communications, Marine and Natural Resources and Others and Kärntner Landesregierung and Others* (CJEU (Grand Chamber), 8 April 2014) <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:62012CJ0293> accessed 10 April 2025.

49 Yannick Alexander Vogel, 'Stretching the Limit, The Functioning of the GDPR's Notion of Consent in the context of Data Intermediary Services' (2022) 8(2) European Data Protection Law Review 238. doi:10.21552/edpl/2022/2/10.

overcome through adaptive legal frameworks that ensure users maintain meaningful control over their data in increasingly complex AI environments. The rapid advancement of AI technologies necessitates specialised legislation that balances innovation with robust rights protection, addressing mounting concerns about fairness, transparency, and accountability in these rapidly evolving systems.

2.3.4. Right to Anonymity

The right to anonymity enables individuals to preserve their identity and personal integrity by preventing unwanted disclosure. This safeguard has become increasingly vital in digital environments where personal information can be exposed and circulated without consent.[50] Traditionally, this right has been understood as an extension of fundamental rights—most notably, the right to privacy[51] and the freedom of expression.[52]

In *Standard Verlagsgesellschaft MBH v. Austria (*App. No 21277/19*)*,[53] the European Court of Human Rights reinforced this principle within the context of online expression. The Court held that compelling the disclosure of anonymous commenters' identities violated Article 10 of the Convention, emphasising that anonymity shields individuals from retaliation and is essential to preserving free and uninhibited participation in democratic debate.

In today's digital age, anonymity has acquired renewed urgency. AI-driven technologies, such as facial recognition, predictive analytics, and pervasive data tracking, have profoundly altered the boundaries of private life. These systems continuously monitor user behaviour, eroding the ability to remain unidentifiable online. As Kettemann et al.[54] observe, anonymity now serves as a crucial safeguard against intrusive surveillance, reinforcing personal agency and autonomy in increasingly datafied societies.

With the proliferation of algorithmic profiling and biometric identification across everyday environments, protecting anonymity is no longer a matter of convenience but a condition for preserving individual freedom, diversity, and democratic discourse in the digital era. Ultimately, safeguarding anonymity is not only about concealing identity but about affirming each individual's right to define their digital presence and personal boundaries in a world of pervasive visibility.

50 Samuel Samiai Andrews, 'Copyright Originality in the Digital Space: The Kingdom of Saudi Arabia's Creatives' in Indranath Gupta (ed), *Handbook on Originality in Copyright: Cases and Materials* (Springer 2023) 1. doi:10.1007/978-981-19-1144-6_9-1.

51 UDHR (n 15) art 12; ICCPR (n 15) art 17.

52 Council of Europe, *European Convention on Human Rights* (Convention for the Protection of Human Rights and Fundamental Freedoms, as amended by Protocols) (ECHR 2013) art 10; ICCPR (n 15) art 19.

53 *Standard Verlagsgesellschaft mbH v Austria (No 3)* App no 21277/19 (ECtHR, 7 December 2021) <https://hudoc.echr.coe.int/fre?i=001-213914> accessed 10 April 2025.

54 Matthias C Kettemann and others, *UNESCO Recommendation on the Ethics of Artificial Intelligence: Conditions for the implementation in Germany* (German Commission for UNESCO 2023).

2.3.5. Right to Be Forgotten

The right to be forgotten reflects one of the most human aspirations in the digital age, the wish to move on from the past and regain control over one's own story. The legal recognition of this right has developed most clearly within the European Union, where it emerged from the long-standing commitment to privacy and data protection. The landmark *Google Spain SL and Google Inc. v. AEPD and Mario Costeja González* (Case C-131/12)[55] captured this idea in a remarkably personal way, when an ordinary citizen sought to erase outdated information that no longer accurately defined who he was. The Court of Justice ruled that individuals may ask search engines to delist links containing outdated or irrelevant personal data, provided this does not override the public's right to know. The decision transformed a private grievance into a principle of digital dignity, later codified in Article 17 of the General Data Protection Regulation (GDPR), which formally grants individuals the right to request the erasure of their personal data.[56]

Later cases, such as *Google LLC v. CNIL* (Case C-507/17),[57] further clarified the boundaries of the right to be forgotten. The Court of Justice affirmed that while individuals deserve the chance to outgrow their digital past, this right cannot extend without limit; it must be balanced with freedom of expression and the public interest. The result is not a promise of invisibility but a nuanced recognition that every person has the right to be more than the sum of their search results.[58]

Yet the challenges today go far beyond search engines. In an era dominated by big data and AI, personal information is not merely stored; it is constantly inferred, replicated, and reassembled by systems that learn from the digital traces individuals leave behind. AI models trained on personal data may reproduce information long after it has been deleted from public sources, raising new questions about whether technological forgetting is even possible.

Still, scholars such as Paolucci[59] emphasise that the right to be forgotten carries profound moral weight; it restores a sense of agency and redemption, allowing individuals to reclaim their narrative from the permanence of the digital archive. In this sense, the right is not only a legal tool but a deeply human one, anchored in dignity, mercy, and the universal need for renewal. It recognises that the digital world—like life itself—should allow space for growth, change, and new beginnings.

55 Case C-131/12 *Google Spain SL and Google Inc v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González* (CJEU (Grand Chamber), 13 May 2014) <https://curia.europa.eu/juris/liste.jsf?num=C-131/12> accessed 10 April 2025.

56 '*Google Spain SL v Agencia Española de Protección de Datos*: Comment Case C-131/12 (May 13, 2014)' (2014) 128(1) *Harvard Law Review* 282.

57 Case C-507/17 *Google LLC, successor in law to Google Inc. v Commission nationale de l'informatique et des libertés (CNIL)* (CJEU, 24 Septembe 2019) <https://curia.europa.eu/juris/liste.jsf?language=en&num=C-507/17> accessed 10 April 2025.

58 Mary Samonte, 'Google v CNIL: The Territorial Scope of the Right to Be Forgotten Under EU Law' (2020) 4(3) European Papers 839. doi:10.15166/2499-8249/332.

59 Paolucci (n 47).

These five digital rights—privacy, internet access, data protection, anonymity, and the right to be forgotten—provide the conceptual framework for evaluating national regulatory approaches. Across global contexts, three findings stand out. First, the *normative core* of human rights remains stable, but their *application* in digital environments varies depending on governance models and technological capabilities. Second, while the European Union leads with legally enforceable standards like the GDPR and AI Act, other regions, including MENA, still rely heavily on ethical or strategic frameworks. Third, the rise of artificial intelligence introduces new tensions between innovation and protection, as data-driven decision-making often exceeds traditional accountability mechanisms. These results highlight the importance of context-specific approaches to digital rights governance. In particular, the MENA region, characterised by rapid technological transformation, strong state involvement, and deeply rooted ethical traditions, offers a distinctive perspective on how universal digital rights are being adapted in practice.

3 DIGITAL RIGHTS AND AI IN THE MENA CONTEXT

3.1. Regional Regulatory Frameworks and Emerging AI Strategies

Building on the conceptual and legal foundation discussed earlier, this section examines how these universal principles translate into regional contexts, particularly in the Middle East and North Africa (MENA) region. Governance approaches to digital rights and AI in the MENA region reflect distinctive regional dynamics and priorities.

Ismail and Ahmad[60] observe that digital transformation across the region has progressed rapidly but unevenly, with Gulf Cooperation Council (GCC) countries generally establishing more advanced governance frameworks than other MENA nations. These differences stem from varying resource availability, institutional capacity, and prioritisation of digital development within national strategies. Several GCC countries have adopted comprehensive AI strategies that engage with both innovation and governance dimensions. The UAE's National Strategy for Artificial Intelligence (2017) was among the first in the region, establishing "responsible AI" as a core pillar alongside economic development goals.[61] Similarly, Qatar's National Artificial Intelligence Strategy emphasises ethical principles including fairness, transparency, and human-centred design.[62]

60 Osama Ismail and Naim Ahmad, 'Ethical and Governance Frameworks for Artificial Intelligence: A Systematic Literature Review' (2025) 19(14) International Journal of Interactive Mobile Technologies121. doi:10.3991/ijim.v19i14.5698.

61 UAE Government, *UAE Strategy for Artificial Intelligence (2017-2031)* (UAE Minister of State for Artificial Intelligence Office 2017) <https://ai.gov.ae/> accessed 10 April 2025.

62 Qatar Ministry of Communications and Information Technology, *National Artificial Intelligence Strategy for Qatar* (MCIT 2019) <https://www.mcit.gov.qa/en/> accessed 10 April 2025.

Many MENA states draw significantly from international frameworks while adapting governance approaches to regionally specific cultural contexts. Qatar's Data Protection Law[63] incorporates GDPR-inspired provisions while reflecting local legal traditions. Similarly, Bahrain's Personal Data Protection Law[64] establishes rights and principles aligned with international standards while maintaining flexibility for national security considerations. The intersection of Islamic legal principles with digital rights frameworks represents a distinctive aspect of regional governance approaches.

Regional cooperation initiatives on digital governance have emerged through bodies including the Arab League and the Gulf Cooperation Council. The Arab Strategy for Information Security and Digital Technologies establish shared principles and cooperation mechanisms, though implementation remains primarily national.[65] As Fatafata and Samaro[66] note, regional initiatives have focused more on cybersecurity cooperation than on comprehensive digital rights frameworks.

Despite these developments, the academic literature on AI ethics and digital rights in the MENA region remains relatively limited compared to the European and North American context.[67] Significant gaps persist in region-specific research on algorithmic fairness, cultural adaptation of AI ethics principles, and implementation studies of digital rights frameworks. This underscores the need for expanded scholarship examining distinctive regional approaches and challenges rather than simply applying external frameworks.

3.2. The Role of Islamic Legal and Ethical Principles

A distinctive feature of AI and digital governance in the MENA context is the interaction between emerging digital rights frameworks and deep-rooted Islamic legal and ethical principles. Ethical foundations derived from *maqāṣid al-sharīʿah*, can enrich regional interpretations of privacy, dignity, justice, and trust as dynamic governance criteria (e.g. *hurmah al-insān*, *adl*, *amānah*). However, integrating these values into AI policy demands more than rhetorical invocation; it demands methodological translation, normative calibration, and institutional embedding.

63 Law of the State of Qatar No 13 of 2016 'On Personal Data Privacy Protection', amended Law No 19 of 2021 <https://www.almeezan.qa/LawView.aspx?opt&LawID=7121&language=ar> accessed 10 April 2025.

64 Law of the Kingdom of Bahrain No 30 of 2018 'On Personal Data Protection' <https://www.bahrain.bh/wps/wcm/connect/ab8b334e-8c6f-4ff9-90b6-94135da559ca/Law+No.+%2830%29+of+2018+DPL.pdf?MOD=AJPERES&CVID=oFapPNI > accessed 10 April 2025.

65 ESCWA and League of Arab States, *Arab Digital Agenda 2023-2033: Arab States Action Programme on Advancing Digital Cooperation and Development* (edn 1.0, UN 2024) <https://www.unescwa.org/publications/arab-digital-agenda-2023-2033> accessed 10 April 2025.

66 Marwa Fatafta and Dima Samaro, *Exposed and Exploited: Data Protection in the Middle East and North Africa* (Access Now 2021) <https://www.skeyesmedia.org/en/News/Reports/29-01-2021/9098> accessed 15 November 2025.

67 Gulf Cooperation Council, *The Guiding Manual on the Ethics of Artificial Intelligence Use in Member States of the Gulf Cooperation Council (GCC)* (Version 1.0, GCC General Secretariat 2023).

Several scholars have explored how Islamic ethics might inform distinctive approaches to emerging challenges such as algorithmic decision-making, data ownership, and privacy by design. In recent research, Ali et al.[68] maintain that Islamic ethics can contribute as a pluralist benchmark for AI evaluation, combining textual sources (*Qur'an, Sunnah*) with *maslahah* (public benefit) and *maqāṣid* reasoning in a dynamic, context-sensitive manner. Elmahjub[69] argues that Islamic ethics should not be considered a static overlay, but rather a living tradition capable of negotiating tensions such as privacy versus utility or fairness versus efficiency through a purpose-based moral methodology.

Islamic jurisprudence provides a set of ethical imperatives that align closely with contemporary human rights standards. Principles such as *hurmah al-insan* (human dignity), *adl* (justice), and *amanah* (trust) resonate strongly with modern values of transparency, accountability, and respect for personal data. As AlKubaisi[70] argues, these principles can complement and reinforce digital rights protections when appropriately integrated into governance frameworks. Trust and responsibility, captured by the concept of *amānah*, are especially important when AI systems are involved in human affairs. From an Islamic perspective, ethical AI should hold developers and users responsible for handling sensitive data and decisions, using tools like audits, clear explanations, and ways to address problems. Western AI ethics are not entirely devoted to this more profound sense of moral accountability; in this regard, Islamic ideas of trusteeship (*amānah*) can offer new ways to embed responsibility into AI systems.

Islamic ethics can make AI governance more meaningful and humane; however, there are still challenges in translating Islamic ethical ideals into a clear regulatory system, especially without a clear, binding legal framework, independent Sharia-based ethical audits, or certified compliance programs. These ideals may not be put into practice.

4 THE INTERSECTION BETWEEN AI GOVERNANCE AND DIGITAL HUMAN RIGHTS IN INTERNATIONAL LEGAL FRAMEWORKS

The intersection between AI governance and digital human rights in international legal frameworks reveals significant variations in regulatory philosophy and implementation strategies. Different jurisdictions have struck varying balances between innovation and rights protection. The European approach generally applies what Jasanoff[71] terms a "precautionary principle," establishing substantial oversight before technologies enter the

68 Ali and others (n 27).

69 Elmahjub (n 26).

70 Abdel Aziz Shaker Hamdan AlKubaisi, 'Ethics of Artificial Intelligence a Purposeful and Foundational Study in Light of the Sunnah of Prophet Muhammad' (2024) 15(11) Religions 1300. doi:10.3390/rel15111300.

71 Sheila Jasanoff, *The Ethics of Invention: Technology and the Human Future* (WW Norton & Company 2016).

market. The American approach has historically emphasised innovation with more limited *ex-ante* restrictions, relying more heavily on market forces and ex-post remedies. Asian approaches often frame AI governance as "enabling innovation" through guidelines and certification rather than comprehensive restrictions.

Yet despite their differences, most jurisdictions face similar challenges when implementing digital rights frameworks. One challenge concerns the elastic nature of concepts such as "fairness" and "transparency," which are subject to varying interpretations, resulting in persistent definitional ambiguity. A second challenge arises from the technical verification difficulties that emerge when assessing complex socio-technical systems that cannot be fully predicted through traditional compliance mechanisms. A third challenge is the "expertise gap" between regulators and the regulated entities they oversee; regulatory bodies often lack the specialised technical capacity to effectively evaluate AI systems for compliance with abstract principles. These issues collectively explain why translating human rights norms into effective AI governance remains an unfinished global project.

4.1. The European AI Act: A Milestone in Human-Rights-Centred Regulation

Against this fragmented background, the European Union's AI Act, which was proposed in 2021 and entered into force on 1 August 2024, stands as the first comprehensive and binding legal framework allocated to managing the dangers and responsibilities of AI.[72] Its central aim is to reconcile technological innovation with the protection of human dignity and fundamental rights.

The Act classifies AI systems according to their potential risk—ranging from prohibited to high, limited, and minimal—with each category carrying corresponding duties relating to transparency, human oversight, and accountability (EU AI Act, Arts. 5–9).[73] For example, the Act prohibits the use of AI for social scoring or real-time biometric surveillance in public spaces, considering these practices incompatible with human dignity. Systems used in law enforcement, healthcare, employment, or education fall into the high-risk category and must undergo strict risk assessments, provide traceable documentation, and ensure that humans remain in control of key decisions.[74] By contrast, tools such as chatbots or emotion-recognition software—classified as limited-risk—are mainly required to inform users that they are interacting with AI, while minimal-risk systems such as spam filters face no additional regulation.[75]

72 Regulation (EU) 2024/1689 (n 12).

73 ibid, arts 5–9.

74 ibid, arts 6(2), 9, 11, 14, 15; annex III (high-risk systems: law enforcement, healthcare, employment, education).

75 ibid, art 50; recital 60 (transparency duties for limited-risk systems; no specific obligations for minimal-risk AI).

This system of categorisation, however, is not without limitations. Some technologies formally labelled as "low risk" may still raise deep ethical concerns, especially when their algorithms amplify bias, manipulate emotions, or operate in an opaque manner.

Despite its ambition, the Act has faced criticism for its ambiguous terminology, reliance on private standard-setting bodies, and limited enforcement pathways. Nevertheless, it remains a historic milestone, transforming abstract principles like privacy, fairness, and non-discrimination into enforceable obligations.

The widespread use of artificial intelligence systems has created a need to adopt mandatory legal measures to protect them, ensure their safety, and guarantee their accountability and adherence to human rights. The lack of binding legislative norms issued by international legislative bodies has left significant gaps in the central part. These gaps in the accountability system reflect a broader challenge of attributing criminal or moral responsibility when algorithmic systems contribute to harmful outcomes. As Abdelaziz[76] argues, the shift from physical to virtual life for human beings has blurred the traditional boundaries of liability, a shift that urgently requires comparative legal systems to rethink how intent, causation, and foreseeability apply to AI-mediated harm.

4.2. Soft-Law Approaches: UNESCO and OECD Frameworks

Beyond the European Union, many international organisations, such as the United Nations, have taken steps to develop ethical baselines for AI safe development and use of artificial intelligence and on the path of the leading world organisations, the Islamic World Educational, Scientific and Cultural Organisation (ICESCO) and the Saudi Data and Artificial Intelligence Authority launched in 2024 the Riyadh Charter for Artificial Intelligence in the Islamic World, which aims to establish an ethical framework for the development and use of artificial intelligence in line with the principles of Islamic moral traditions.

The UNESCO Recommendation on the Ethical Use of Artificial Intelligence, adopted by 193 member countries in November 2021, is particularly significant.[77] Unlike binding regulatory instruments, which rely on prohibitions and strict compliance measures, the Recommendation adopts a life-cycle approach that emphasises human dignity, the protection of individuals' privacy, and the inclusion of people in the development of AI. It challenges states to embed human rights into AI governance and seeks to ensure that technological innovation strengthens societies, respects liberty, and unites individuals. Its holistic approach sets out ten core principles—proportionality, safety and security, privacy and data protection, governance, inclusivity, responsibility and accountability, transparency and explainability, sustainability, human oversight, public AI literacy, and fairness and unbiased AI.

76 Dalia Kadry Ahmed Abdelaziz, 'Incitement to Suicide in the Digital Age: A Comparative Legal Study of Criminal Liability' (2025) 5(6) Journal of Posthumanism 684. doi:10.63332/joph.v5i6.2105

77 UNESCO (n 10).

Similarly, the OECD AI guidelines offer a set of recommendations with clear aims for the responsible stewardship of the development and use of AI that respects and protects human rights and includes democratic rights and core freedoms.[78] These principles were adopted by OECD member countries and remain a significant step towards the formation of an AI regulatory framework. The OECD AI Principles share significant similarities with UNESCO's principles regarding inclusiveness, collaboration, openness, and responsibility.[79] The principles also provide that, to ensure fair and balanced rights for individuals, the AI systems used to process and reach decisions that affect people should be transparent and explainable to all stakeholders involved.

5 DIGITAL HUMAN RIGHTS WITHIN THE SAUDI ARABIA'S LEGAL FRAMEWORK: A PRELIMINARY ASSESSMENT

The Kingdom of Saudi Arabia has developed a multi-layered legal framework governing digital activities and data protection, comprising the Personal Data Protection Law (PDPL, 2021, amended 2023), the Anti-Cyber Crime Law (2007), and various sector-specific regulations. This framework operates within the broader context of Vision 2030, which explicitly prioritises digital transformation while recognising the need for rights protections.[80]

The regulatory architecture reflects a hybrid approach: comprehensive data protection provisions modelled on international standards[81] (particularly GDPR-inspired elements in the PDPL) coexist with cybersecurity legislation predating the AI era. The Saudi Data and Artificial Intelligence Authority (SDAIA) has issued AI Ethics Principles (2023), though these remain non-binding guidance rather than enforceable law. The Saudi Data and Privacy Protection Authority (SDPPA) serves as the primary enforcement body for data protection.[82]

This section evaluates Saudi Arabia's framework against the international benchmarks established in Section 3, examining both areas of alignment with global standards and regulatory gaps that limit comprehensive digital rights protection in AI-driven contexts. The analysis begins by assessing the achievements of the PDPL, before examining its limitations, and before turning to the contemporary relevance of the Anti-Cybercrime Law. The section concludes with a comparative assessment of the Kingdom's overall positioning within international governance models.

78 OECD (n 9); 'OECD AI Principles' (*OECD*, 2025) <https://www.oecd.org/en/topics/ai-principles.html> accessed 6 October 2025.

79 Nicholas Kluge Corrêa and others, 'Worldwide AI Ethics: A Review of 200 Guidelines and Recommendations for AI Governance' (2023) 4(10) Patterns 100857. doi:10.1016/j.patter.2023.100857.

80 Mohammad Rashed Albous, Odeh Rashed Al-Jayyousi and Melodena Stephens, 'AI Governance in the GCC States: A Comparative Analysis of National AI Strategies' (2025) 82 Journal of Artificial Intelligence Research 2389. doi:10.1613/jair.1.17619.

81 Bendary and Rajadurai (n 21).

82 Ibrahim (n 14).

5.1. Personal Data Protection Law (PDPL)

The Personal Data Protection Law (PDPL), enacted on 16 September 2021 and amended on 27 March 2023,[83] constitutes Saudi Arabia's primary legal framework for data protection. Enforced by the Saudi Data and Privacy Protection Authority (SDPPA), the PDPL establishes comprehensive provisions governing the collection, processing, and storage of personal information, demonstrating the Kingdom's recognition of data privacy as an essential component of digital rights protection.

The PDPL reflects influence from established international frameworks, particularly the Council of Europe's Convention 108[84] and the EU's General Data Protection Regulation (GDPR).[85] The law incorporates several internationally recognised principles, including purpose limitation, data minimisation, and consent requirements for data processing. Article 5 mandates explicit consent for data processing, while Article 12 imposes transparency requirements for data use. Article 19 enforces data minimisation principles, requiring data collection to be limited to essential information.[86] These provisions constitute the foundational legal framework for digital human rights in Saudi Arabia, particularly regarding the right to informational self-determination that enables individuals to exercise control over their personal information.[87]

The Kingdom has successfully implemented additional protective measures that mirror global best practices. Article 24 requires data breach notification to both authorities and affected individuals without undue delay—a provision consistent with leading international data protection regulations. Article 32 mandates the appointment of Data Protection Officers within organisations that process personal data, and establishes an internal compliance mechanism.[88] The establishment of the SDPPA as a dedicated regulatory body reflects international trends toward specialised data protection oversight, providing institutional capacity for enforcement and guidance. The law grants individuals specific rights regarding their personal data, including access, correction, and deletion rights, demonstrating the Kingdom's commitment to empowering citizens with meaningful

83 Royal Decree of the Kingdom of Saudi Arabia No M/19 of 09/02/1443 AH (n 14).

84 Council of Europe, *Convention 108 + : Convention for the Protection of Individuals with Regard to the Processing of Personal Data* (CoE 2018) <https://www.coe.int/en/web/data-protection/convention108-and-protocol> accessed 6 October 2025.

85 Regulation (EU) 2016/679 (n 46).

86 Mutaz Abdulaziz Alkhedhairy, 'Balancing Privacy and Risk: A Critical Analysis of Personal Data Use as Governed by Saudi Insurance Law' (2025) 14(4) Laws 47. doi:10.3390/laws14040047.

87 Nick O'Connell, 'An overview of Saudi Arabia's new Personal Data Protection Law' (*Al Tamimi & Co*, September 2021) <https://www.tamimi.com/law-update-articles/an-overview-of-saudi-arabias-new-personal-data-protection-law/> accessed 15 November 2025.

88 Marianne Rahme, 'Data Protection in Saudi Arabia: Comparative Analysis General Data Protection Regulation Kingdom of Saudi Arabia KSA' (*SMEX*, 10 February 2022) <https://smex.org/data-protection-in-saudi-arabia-comparative-analysis/> accessed 15 November 2025.

control over their personal information. Substantial penalties for non-compliance—potentially reaching 5 million Saudi Riyals (approximately 1.3 million USD) for serious violations—signal strong enforcement commitment.[89]

Despite these achievements, several areas reveal divergence from international best practices and warrant additional attention to address contemporary technological challenges. Article 16 delineates multiple circumstances that permit the waiver of consent requirements, including public interest, vital interests protection, national security considerations, credit referencing, and research purposes. These exceptions, though providing operational flexibility, create broader latitude than comparable provisions in the GDPR, which may weaken the foundational principle of informed consent that underpins modern data protection regimes.

The current legal framework provides limited algorithmic accountability provisions, creating gaps in oversight of AI decision-making systems that increasingly affect citizens' lives across sectors, from employment to public services. The law currently lacks specific provisions addressing emerging technological applications. Spatiotemporal data and AI-generated information receive limited attention despite their widespread use across healthcare, transportation, and smart city initiatives. Post-mortem data processing, particularly regarding facial recognition and predictive analytics, lacks clear regulatory guidelines. Additionally, the framework does not explicitly address bias mitigation in automated systems or transparency requirements for AI decision-making processes, creating a significant regulatory gap, as the Kingdom lacks unified legislation that addresses the full spectrum of AI-related rights and risks.[90]

These gaps create challenges for comprehensive digital rights protection as artificial intelligence and machine learning technologies expand throughout Saudi society. Developing provisions for algorithmic transparency and accountability would strengthen the framework's capacity to address automated decision-making in employment, judicial processes, and public services. As the Kingdom advances its digital transformation agenda under Vision 2030, evolving the PDPL to encompass these emerging areas would enhance alignment with international best practices while supporting technological innovation objectives.

89 'Penalties for Non-Compliance with PDPL' (*Standard Touch*, 2025) <https://standardtouch.com/pdpl-penalties-saudi-arabia/> accessed 6 October 2025.

90 Adamantia Rachovitsa, 'Engineering and Lawyering Privacy by Design: Understanding Online Privacy Both as a Technical and An International Human Rights Issue' (2016) 24(4) *International Journal of Law and Information Technology* 374. doi:10.1093/ijlit/eaw012.

5.2. 2007 Anti-Cybercrime Law and Digital Rights Implications

The 2007 Anti-Cybercrime Law[91] complements the PDPL but predates both contemporary data protection standards and the proliferation of AI technologies. Article 3(1) prohibits unauthorised data interception through information networks. Article 3(4) addresses privacy invasion through mobile devices and similar technologies.[92] These provisions establish baseline privacy protections but do not explicitly address AI-powered surveillance technologies such as facial recognition systems or predictive analytics.[93]

Article 6(1) regulates content production and transmission that impinges on public order, religious values, or privacy. Article 7(2) addresses unauthorised system access that may result in the obtaining of data relevant to national security or economic interests. These provisions provide flexibility for addressing evolving threats but lack specific guidance on algorithmic content moderation or AI-driven security measures. Such systems currently operate throughout the Kingdom, including biometric identification at border entry points, the ABSHER digital government services platform,[94] and AI-powered urban management technologies deployed in NEOM smart city initiatives.[95]

Article 14 designates the Communications and Information Technology Commission to provide technical support to security agencies during investigations. The law does not address algorithmic bias in automated enforcement systems, transparency requirements for AI-powered investigative tools, or citizens' rights to challenge automated decisions.[96] International courts have increasingly recognised these gaps as human rights concerns. The European Court of Human Rights found that mass surveillance systems lacking adequate safeguards violate privacy rights in *Big Brother Watch and Others v. United Kingdom* (Apps. Nos 58170/13, 62322/14 and 24960/15).[97] The Court of Justice of the European Union invalidated data transfer mechanisms where surveillance frameworks

91 Royal Decree of the Kingdom of Saudi Arabia No M/17 of 8 Rabi'I 1428H 'Anti-Cyber Crime Law' (26 March 2007) <https://www.wipo.int/wipolex/en/legislation/details/14570> accessed 10 April 2025.

92 Selma Dilek, Hüseyin Çakır and Mustafa Aydın, 'Applications of Artificial Intelligence Techniques to Combating Cybercrimes: A Review' (*arXiv preprint*, 12 February 2015) arXiv:1502.03552. doi:10.48550/arXiv.1502.03552.

93 Thomas C King and others, 'Artificial Intelligence Crime: An Interdisciplinary Analysis for Foreseeable Threats and Solutions' (2020) 26 Science and Engineering Ethics 89. doi:10.1007/s11948-018-00081-0.

94 Ministry of Interior of the Kingdom of Saudi Arabia, *Absher Platform* (2025) <https://www.absher.sa/portal/landing.html> accessed 6 October 2025.

95 NEOM, 'Technology and Digital' (*NEOM Official Website*, 2025) <https://www.neom.com> acsessed 6 October 2025.

96 Cristos Velasco, 'Cybercrime and Artificial Intelligence. An Overview of the Work of International Organization on Criminal Justice and the International Applicable Instruments' (2022) 23 ERA Forum 109. doi:10.1007/s12027-022-00702-z.

97 *Big Brother Watch and Others v United Kingdom* Apps nos 58170/13, 62322/14 and 24960/15 (ECtHR, 25 May 2021) <https://hudoc.echr.coe.int/fre?i=001-210077> accessed 6 October 2025.

provided insufficient individual protections.[98] The CJEU also recognised individuals' rights to request deletion of inadequate or outdated personal data, a principle not yet incorporated into Saudi legislation.[99]

The 17-year gap between the law's enactment and current AI capabilities underscores the need for updated provisions that address algorithmic accountability and transparency. Developing such provisions would strengthen the Kingdom's framework capacity to balance security objectives with digital rights protections as AI technologies expand throughout Saudi society.

6 CONCLUSION AND RECOMMENDATIONS

K.S.A has established foundational digital governance structures that demonstrate general alignment with international standards, particularly through the Personal Data Protection Law (2021, amended 2023) and the creation of specialised regulatory institutions. The PDPL incorporates core data protection principles, including purpose limitation, data minimisation, consent requirements, and breach notification, while granting individuals meaningful rights over their personal information. These achievements position the Kingdom comparably to other jurisdictions pursuing digital transformation under comprehensive data protection regimes.

However, the accelerating deployment of AI systems across governance, healthcare, education, and smart city infrastructure has outpaced the legal framework's capacity to address algorithmic decision-making. Three critical gaps persist: the PDPL does not recognise data protection as a fundamental human right; neither the PDPL nor the 2007 Anti-Cybercrime Law adequately addresses AI-specific challenges, including algorithmic accountability, bias mitigation, and transparency requirements; and fragmented regulatory oversight leaves individuals without clear mechanisms to challenge automated decisions affecting their interests.

The path forward requires targeted legal reforms. Amending the PDPL to explicitly recognise data protection as a fundamental right would strengthen its constitutional foundation. Comprehensive AI governance legislation that consolidates SDAIA's ethical guidelines into binding requirements would establish clear obligations for high-risk applications, mandatory transparency standards, and enforceable bias-prevention measures. Expanding the PDPL's scope to explicitly cover AI-generated data, biometric information, and spatiotemporal analytics would address technological developments since

98 Case C-311/18 *Data Protection Commissioner v Facebook Ireland Limited and Maximillian Schrems* (CJEU (Grand Chamber), 16 July 2020) <https://curia.europa.eu/juris/liste.jsf?num=C-311/18> accessed 6 October 2025.

99 Case C-131/12 *Google Spain SL and Google Inc v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González* (n 55).

the law's enactment. Strengthening enforcement mechanisms—including accessible complaint procedures, rights to explanation for automated decisions, and meaningful remedies for violations—would transform abstract protections into actionable rights.

KSA's distinctive position presents both challenge and opportunity. The Kingdom is undergoing accelerated digital transformation while simultaneously integrating Islamic ethical principles into its legal system. Operationalising concepts of human dignity (*hurmah al-insan*), justice (*adl*), and trust (*amanah*) within technical standards and compliance frameworks would position the KSA model for governance approaches that harmonise international norms with cultural values. Rather than viewing Vision 2030's technological ambitions and robust rights protections as competing imperatives, the Kingdom can demonstrate their mutual reinforcement—innovation flourishes most sustainably within frameworks that safeguard human dignity and accountability.

By addressing these regulatory gaps through comprehensive legal reform, Saudi Arabia can establish itself as a regional leader in rights-based AI governance, demonstrating that rapid technological advancement and fundamental rights protection are not opposing forces but complementary pillars of sustainable digital transformation.

REFERENCES

1. Abdelaziz DKA, 'Incitement to Suicide in the Digital Age: A Comparative Legal Study of Criminal Liability' (2025) 5(6) Journal of Posthumanism 684. doi:10.63332/joph.v5i6.2105
2. Albous MR, Al-Jayyousi OR and Stephens M, 'AI Governance in the GCC States: A Comparative Analysis of National AI Strategies' (2025) 82 Journal of Artificial Intelligence Research 2389. doi:10.1613/jair.1.17619
3. Alhejaili MOM, 'Integrating Smart Contracts into the Legal Framework of Saudi Arabia' (2025) 67(2) International Journal of Law and Management 230. doi:10.1108/IJLMA-03-2024-0086
4. Alhejaili MOM, 'Securing the Kingdom's e-Commerce Frontier: Evaluation of Saudi Arabia's Cybersecurity Legal Frameworks' (2024) 13(2) Journal of Governance & Regulation 275. doi:10.22495/jgrv13i2siart4
5. Ali F and others, 'Islamic Ethics and AI: An Evaluation of Existing Approaches to AI using Trusteeship Ethics' (2025) 38(2) Philosophy & Technology 120. doi:10.1007/s13347-025-00922-4
6. Alkhedhairy MA, 'Balancing Privacy and Risk: A Critical Analysis of Personal Data Use as Governed by Saudi Insurance Law' (2025) 14(4) Laws 47. doi:10.3390/laws14040047
7. AlKubaisi AASH, 'Ethics of Artificial Intelligence a Purposeful and Foundational Study in Light of the Sunnah of Prophet Muhammad' (2024) 15(11) Religions 1300. doi:10.3390/rel15111300

8. AllahRakha N, 'UNESCO's AI Ethics Principles: Challenges and Opportunities' (2024) 2(9) *International Journal of Law and Policy* 24. doi:10.59022/ijlp.225

9. Alzahrani RB, 'An Overview of AI Data Protection in the Context of Saudi Arabia' (2024) 3(3) International Journal for Scientific Research 199. doi:10.59992/IJSR.2024.v3n3p8 [in Arabic]

10. Andrews SS, 'Copyright Originality in the Digital Space: The Kingdom of Saudi Arabia's Creatives' in Gupta I (ed), *Handbook on Originality in Copyright: Cases and Materials* (Springer 2023) 1. doi:10.1007/978-981-19-1144-6_9-1

11. Bendary MG and Rajadurai J, 'Emerging Technologies and Public Innovation in the Saudi Public Sector: An Analysis of Adoption and Challenges Amidst Vision 2030' (2024) 29(1) The Innovation Journal: The Public Sector Innovation Journal 1

12. Cerf V, 'Internet Access is Not a Human Right' *The New York Times* (New York, 4 January 2012) A25

13. Cong W, 'Understanding Human Rights on the Internet: An Exercise of Translation?' (2017) 22(1-2) Tilburg Law Review 138. doi:10.1163/22112596-02201007

14. Custers B, 'New Digital Rights: Imagining Additional Fundamental Rights for the Digital Era' (2022) 44 *Computer Law & Security Review* 105636. doi:10.1016/j.clsr.2021.105636

15. De Hert P and Kloza D, 'Internet (access) as a New Fundamental Right: Inflating the Current Rights Framework?' (2012) 3(3) European Journal of Law and Technology 1

16. Dilek S, Çakır H and Aydın M, 'Applications of artificial intelligence techniques to combating cybercrimes: A review' (*arXiv preprint,* 12 February 2015) arXiv:1502.03552. doi:10.48550/arXiv.1502.03552

17. Elmahjub E, 'Artificial Intelligence (AI) in Islamic Ethics: Towards Pluralist Ethical Benchmarking for AI' (2023) 36(4) Philosophy & Technology 73. doi:10.1007/s13347-023-00668-x

18. Fatafta M and Samaro D, *Exposed and Exploited: Data Protection in the Middle East and North Africa* (Access Now 2021)

19. Fjeld G and others, *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI* (Berkman Klein Center for Internet & Society Research Publication Series no 2020-1, Harvard University 2020). doi:10.2139/ssrn.3518482

20. Hildebrandt M, *Smart Technologies and the End(s) of Law* (Edward Elgar 2015)

21. Ibrahim NMH, 'Artificial Intelligence (AI) and Saudi Arabia's Governance' (2024) 40(4) Journal of Developing Societies 500. doi:10.1177/0169796X241288590

22. Ismail O and Ahmad N, 'Ethical and Governance Frameworks for Artificial Intelligence: A Systematic Literature Review' (2025) 19(14) International Journal of Interactive Mobile Technologies121. doi:10.3991/ijim.v19i14.5698

23. Jasanoff S, *The Ethics of Invention: Technology and the Human Future* (WW Norton & Company 2016)

24. Jobin A, Ienca M and Vayena E, 'The Global Landscape of AI Ethics Guidelines' (2019) 1(9) Nature Machine Intelligence 389. doi:10.1038/s42256-019-0088-2

25. Jørgensen RF, *Framing the Net: The Internet and Human Rights* (Edward Elgar 2013). doi:10.4337/9781782540809

26. Kettemann MC and others, *UNESCO Recommendation on the Ethics of Artificial Intelligence: Conditions for the implementation in Germany* (German Commission for UNESCO 2023)

27. King TC and others, 'Artificial Intelligence Crime: An Interdisciplinary Analysis for Foreseeable Threats and Solutions' (2020) 26 Science and Engineering Ethics 89. doi:10.1007/s11948-018-00081-0

28. Kluge Corrêa N and others, 'Worldwide AI Ethics: A Review of 200 Guidelines and Recommendations for AI Governance' (2023) 4(10) Patterns 100857. doi:10.1016/j.patter.2023.100857

29. Kostenko OM and others, '"Legal Personality" of Artificial Intelligence: Methodological Problems of Scientific Reasoning by Ukrainian and EU Experts' (2023) 39(4) AI and Society 1683. doi:10.1007/s00146-023-01641-0

30. Land MK and Aronson JD (eds), *New Technologies for Human Rights Law and Practice* (CUP 2018) doi:10.1017/9781316838952

31. Larsson S and Heintz F, 'Transparency in Artificial Intelligence' (2020) 9(2) Internet Policy Review. doi:10.14763/2020.2.1469

32. Mathiesen K, 'The Human Right to Internet Access: A Philosophical Defense' (2012) 18 The International Review of Information Ethics 9. doi:10.29173/irie299

33. O'Connell N, 'An overview of Saudi Arabia's new Personal Data Protection Law' (*Al Tamimi & Co*, September 2021) <https://www.tamimi.com/law-update-articles/an-overview-of-saudi-arabias-new-personal-data-protection-law/> accessed 15 November 2025.

34. Pangrazio L and Sefton-Green J, 'Digital Rights, Digital Citizenship and Digital Literacy: What's the Difference?' (2021) 10(1) *Journal of New Approaches in Educational Research* 15. doi:10.7821/naer.2021.1.616

35. Paolucci F, 'Enhancing Oversight and Addressing Gaps: Assessing the Impact of the AI Act on Biometric Identification Systems' in Menéndez González N and Mobilio G (eds), *Next Democratic Frontiers for Facial Recognition Technology (FRT): The Legal, Ethical and Democratic Implications of FRT* (Springer 2025) 71. doi:10.1007/978-3-031-89794-8_5

36. Penney J, 'Open Connectivity, Open Data: Two Dimensions of the Freedom to Seek, Receive and Impart Information in the New Zealand Bill of Rights' (2012) 4 Victoria University of Wellington Law Review: Working Paper Series 1

37. Rachovitsa A, 'Engineering and Lawyering Privacy by Design: Understanding Online Privacy Both as a Technical and An International Human Rights Issue' (2016) 24(4) *International Journal of Law and Information Technology* 374. doi:10.1093/ijlit/eaw012

38. Rahme M, 'Data Protection in Saudi Arabia: Comparative Analysis General Data Protection Regulation Kingdom of Saudi Arabia KSA' (*SMEX*, 10 February 2022) <https://smex.org/data-protection-in-saudi-arabia-comparative-analysis/> accessed 15 November 2025.

39. Reidenberg JR, 'Lex Informatica: The Formulation of Information Policy Rules Through Technology' (1997) 76(3) Texas Law Review 553

40. Samonte M, 'Google v CNIL: The Territorial Scope of the Right to Be Forgotten Under EU Law' (2020) 4(3) European Papers 839. doi:10.15166/2499-8249/332

41. Schwartz A, 'Chicago's Video Surveillance Cameras: A Pervasive and Poorly Regulated Threat to Our Privacy' (2013) 11(2) Northwestern Journal of Technology and Intellectual Property 47.

42. Tully S, 'A Human Right to Access the Internet? Problems and Prospects' (2014) 14(2) Human Rights Law Review 175. doi:10.1093/hrlr/ngu011

43. Veale M and Zuiderveen Borgesius F, 'Demystifying the Draft EU Artificial Intelligence Act' (2021) 22(4) Computer Law Review International 97. doi:10.9785/cri-2021-220402

44. Velasco C, 'Cybercrime and Artificial Intelligence. An Overview of the Work of International Organization on Criminal Justice and the International Applicable Instruments' (2022) 23 ERA Forum 109. doi:10.1007/s12027-022-00702-z

45. Vogel YA, 'Stretching the Limit, The Functioning of the GDPR's Notion of Consent in the context of Data Intermediary Services' (2022) 8(2) European Data Protection Law Review 238. doi:10.21552/edpl/2022/2/10

46. Winfield AFT and Jirotka M, 'Ethical Governance is Essential to Building Trust in Robotics and Artificial Intelligence Systems' (2018) 376(2133) Philosophical Transactions of the Royal Society A 20180085. doi:10.1098/rsta.2018.0085

AUTHORS INFORMATION

Soumaya Khammassi
PhD (Law), Assistant Professor, College of Law, Prince Sultan University, Riyadh, Saudia Arabia
skhammassi@psu.edu.sa
https://orcid.org/0009-0008-7810-0032
Co-author, responsible for conceptualization, methodology, investigation, writing – original draft and writing – review & editing.

Yusra AlShanqityi*
PhD (Law), Assistant Professor, College of Law, Prince Sultan University, Riyadh, Saudia Arabia
yshanqityi@psu.edu.sa
https://orcid.org/0009-0004-9275-2046
Corresponding author, responsible for conceptualization, methodology, investigation and writing – review & editing.

Competing interests: No competing interests were disclosed. Any potential conflict of interest must be disclosed by authors.

Disclaimer: The authors declare that their opinion and views expressed in this manuscript are free of any impact of any organizations.

RIGHTS AND PERMISSIONS

ADDITIONAL INFORMATION

The authors are thankful to the Governance and Policy Design Research Lab (GPDRL) and to Prince Sultan University for their academic support and for providing APC for this publication.

EDITORS

ABOUT THIS ARTICLE

Cite this article
Khammassi S and AlShanqityi Y, 'Digital Rights, AI, and the Law: International Perspectives on Saudi Arabia's Legal Framework and International Experience' (2025) 8(Spec) Access to Justice in Eastern Europe 146-77 <https://doi.org/10.33327/AJEE-18-8.S-a000154>

DOI: https://doi.org/10.33327/AJEE-18-8.S-a000154

Keywords: *digital rights, artificial intelligence, human rights, legislation, data protection, cybersecurity, privacy, e-government, Vision 2030.*

DETAILS FOR PUBLICATION

Date of submission: 16 Apr 2025
Date of acceptance: 08 Sep 2025
Online First publication: 08 Dec 2025
Last Published: 30 Dec 2025
Whether the manuscript was fast tracked? - No
Number of reviewer report submitted in first round: 3 reports
Number of revision rounds: 1 round with conditionally acceptance

Technical tools were used in the editorial process:
Plagiarism checks - Turnitin from iThenticate https://www.turnitin.com/products/ithenticate/
Scholastica for Peer Review https://scholasticahq.com/law-reviews

AI DISCLOSURE STATEMENT

The corresponding author confirmed that the manuscript was written by the authors. AI tools (Claude and Grammarly) were used for spelling, grammar, stylistic refinement, and citation format verification. No generative AI was used to create original content, research ideas, or legal analyses.

АНОТАЦІЯ УКРАЇНСЬКОЮ МОВОЮ

Дослідницька стаття

ЦИФРОВІ ПРАВА, ШТУЧНИЙ ІНТЕЛЕКТ ТА ЗАКОНОДАВСТВО: МІЖНАРОДНІ ПЕРСПЕКТИВИ ЩОДО ПРАВОВОЇ СИСТЕМИ САУДІВСЬКОЇ АРАВІЇ Т А МІЖНАРОДНИЙ ДОСВІД

Сумая Хаммассі та Юсра Аль-Шанкіті*

АНОТАЦІЯ

Вступ. *У цьому дослідженні аналізується глобальна еволюція цифрових прав та управління штучним інтелектом, а також розглядаються наслідки для правової системи Саудівської Аравії. Оскільки штучний інтелект все більше впроваджується в повсякденне життя, існує нагальна потреба оцінити його вплив на основоположні права людини, особливо з огляду на відсутність глобального консенсусу щодо визначення та сфери застосування «цифрових прав людини». Ініціатива Королівства Саудівської Аравії (KCA) «Vision 2030» надає унікальну можливість розробити рамки управління ШІ, які відповідатимуть міжнародним стандартам і водночас відображатимуть місцеві культурні цінності та ісламські етичні принципи.*

Методи. *У цьому дослідженні використовується якісний аналітичний підхід для вивчення чинної правової бази Саудівської Аравії, яка регулює ШІ та цифрові права людини. Методологія передбачає комплексний аналіз законодавства Саудівської Аравії, зокрема Закону про захист персональних даних (PDPL) та етичних принципів Управління з питань даних та штучного інтелекту Саудівської Аравії (SDAIA), у порівнянні з міжнародними правовими інструментами, включно з Рекомендаціями ЮНЕСКО щодо етики штучного інтелекту, Законом ЄС про штучний інтелект та Принципами ОЕСР з питань штучного інтелекту. У статті оцінюється відповідність міжнародним стандартам та ефективність у вирішенні нових проблем, що стосуються цифрових прав в епоху ШІ.*

Результати та висновки. *Дослідження показує, що хоча Саудівська Аравія досягла прогресу завдяки структурам PDPL та SDAIA, все ж значні нормативні прогалини*

залишаються. PDPL має обмеження у вирішенні сучасних проблем у сфері штучного інтелекту, зокрема алгоритмічну підзвітність, зменшення упередженості та комплексний захист даних у контексті штучного інтелекту. У результаті дослідження було виявлено критичні недоліки, зокрема занадто широкі винятки щодо вимог стосовно згоди, недостатньо чіткі положення щодо прозорості алгоритмів і фрагментований регулятивний нагляд. Рекомендації передбачають створення спеціалізованих органів нагляду, розробку етичних рамок, адаптованих до контексту Саудівської Аравії, і більшого залучення експертів до ухвалення рішень, пов'язаних з управлінням ШІ. Ця стаття робить свій внесок, пропонуючи порівняльну оцінку системи цифрових прав Саудівської Аравії з провідними міжнародними інструментами, висвітлюючи реформи, необхідні для управління, заснованого на культурних традиціях та узгодженого на глобальному рівні.

Ключові слова: *цифрові права, штучний інтелект, права людини, законодавство, захист даних, кібербезпека, конфіденційність, електронне урядування, «Vision 2030».*

ABSTRACT IN ARABIC

مقال بحثي

الحقوق الرقمية والذكاء الاصطناعي والقانون: منظور دولي لإطار السعودية القانونيوالخبرة الدولية

سمية الخماسي ويسرى الشنقيطي*

الملخص

الخلفية: *يقدّم هذا البحث تحليلًا للتطوّر العالمي للحقوق الرقمية وحوكمة الذكاء الاصطناعي، ويبحث في انعكاساته على الإطار القانوني في المملكة العربية السعودية. ومع تزايد حضور الذكاء الاصطناعي في مختلف جوانب الحياة اليومية، تبرز الحاجة الملحّة إلى تقييم أثره على حقوق الإنسان الأساسية، خصوصًا في ظل غياب توافق دولي حول تعريف ومجال «الحقوق الرقمية للإنسان». وتقدّم مبادرة رؤية المملكة 2030 فرصة فريدة لتطوير أطر حوكمة للذكاء الاصطناعي تتماشى مع المعايير الدولية، مع مراعاة القيم الثقافية المحلية والمبادئ الأخلاقية الإسلامية.*

المنهجية: *يعتمد هذا البحث مقاربة تحليلية نوعية لفحص الإطار القانوني الحالي في المملكة العربية السعودية المنظّم للذكاء الاصطناعي والحقوق الرقمية للإنسان. وتشمل المنهجية تحليلًا تشريعيًا شاملًا*

للأنظمة السعودية، وعلى وجه الخصوص نظام حماية البيانات الشخصية (PDPL) وإرشادات الأخلاقيات الصادرة عن الهيئة السعودية للبيانات والذكاء الاصطناعي (SDAIA)، مع مقارنتها بالأدوات القانونية الدولية، بما في ذلك توصية اليونسكو لأخلاقيات الذكاء الاصطناعي، وقانون الاتحاد الأوروبي للذكاء الاصطناعي، وإرشادات منظمة التعاون الاقتصادي والتنمية (OECD). ويقيّم البحث مدى الاتساق مع المعايير الدولية وفعالية هذه الأطر في معالجة تحديات الحقوق الرقمية في عصر الذكاء الاصطناعي.

النتائج والاستنتاجات: *يكشف البحث أنه على الرغم من أن المملكة العربية السعودية حققت تقدمًا ملحوظًا من خلال نظام حماية البيانات الشخصية وأطر إرشادات الأخلاقيات الصادرة عن الهيئة السعودية للبيانات والذكاء الاصطناعي، فإن فجوات تنظيمية كبيرة ما تزال قائمة. ويُظهر نظام حماية البيانات الشخصية جوانب قصور في التعامل مع تحديات الذكاء الاصطناعي المعاصرة، بما في ذلك مساءلة الخوارزميات، والحد من التحيّز، والحماية الشاملة للبيانات في سياقات الذكاء الاصطناعي. كما يحدّد البحث أوجه قصور أساسية، منها الاستثناءات الواسعة لمتطلبات الموافقة، وضعف الأحكام المتعلقة بشفافية الخوارزميات، وتشتّت الجهات الرقابية. وتشمل التوصيات إنشاء هيئات رقابية متخصصة، وتطوير أطر أخلاقية تراعي السياق السعودي، وزيادة إشراك الخبراء في القرارات المتعلقة بحوكمة الذكاء الاصطناعي. وتُسهم هذه الورقة في تقديم تقييم مقارن لإطار الحقوق الرقمية في المملكة العربية السعودية مقابل أبرز الأدوات الدولية، مع تسليط الضوء على الإصلاحات اللازمة لضمان حوكمة متجذّرة ثقافيًا ومتوافقة عالميًا.*

Access to Justice in Eastern Europe
ISSN 2663-0575 (Print)
ISSN 2663-0583 (Online)
http://ajee-journal.com

Research Article

THE ROLE OF STATUTORY LAW IN REGULATING ARTIFICIAL INTELLIGENCE: BALANCING INNOVATION AND RESPONSIBILITY

Abdesselam Salmi, Bhupal Bhattacharya, Sarmistha Bhattacharya and Tarek Abo El Wafa*

DOI:
https://doi.org/10.33327/AJEE-18-8.S-a000150

Date of submission: 24 Apr 2025
Date of acceptance: 17 Jul 2025
Online First publication: 04 Dec 2025
Last Published: 30 Dec 2025

ABSTRACT

Background: *Artificial Intelligence (AI) poses profound governance challenges, as its rapid integration across critical sectors exacerbates risks of discrimination, privacy violations, and accountability gaps. Statutory law, which traditionally underpins national legal systems, is proving increasingly insufficient to regulate the ethical, social, and economic implications of AI. Its structural rigidity, coupled with lengthy legislative processes and jurisdictional fragmentation, renders it ill-equipped to respond to the fast-evolving nature of algorithmic technologies. Consequently, regulatory gaps emerge in high-risk applications such as predictive policing, biometric surveillance, medical diagnostics, and autonomous weapons domains, where errors or biases can lead to irreversible harm. Many existing legal norms were crafted without anticipating the complexity and opacity of machine learning systems, including their potential to operate in ways that defy traditional notions of human intention, liability, and foreseeability. As a result,*

there is an urgent need for scholarly engagement with the conceptual and practical tensions between innovation and regulation in the AI context. This includes exploring adaptive legal frameworks, hybrid governance models, and the integration of ethical principles into technological design.

Methods: *This study employs a comparative legal analysis of AI regulatory frameworks across key jurisdictions (EU, US, China, Brazil, UK), combined with doctrinal research of legislative texts and case law. The methodology integrates a systematic review of primary sources (e.g., EU AI Act, US Algorithmic Accountability Act drafts, China's GenAI Interim Measures), a qualitative assessment of secondary literature and institutional reports, application of the Issue-Rule-Application-Conclusion framework to evaluate regulatory efficacy, and a cross-jurisdictional examination of enforcement mechanisms and liability standards.*

Results and conclusions: *The analysis reveals statutory law's critical limitations, jurisdictional divergences in risk classification (e.g., the EU's ex-ante conformity assessments vs. the US's sectoral ex-post enforcement), liability fragmentation, and enforcement gaps. Crucially, statutory approaches alone cannot balance innovation promotion with ethical constraints: excessive regulation stifles R&D, while lax frameworks enable societal harm. The study concludes that effective governance requires complementary ethical frameworks that embed transparency, bias auditing, and human oversight; international harmonisation of liability standards and risk protocols; adaptive regulatory sandboxes for real-world testing; and multistakeholder collaboration to design context-sensitive implementations.*

1 INTRODUCTION

As artificial intelligence (AI) advances and its applications expand, its significance continues to grow, exerting an increasing impact on society and shaping future development.[1] The majority of economic sectors, social interactions, and technical breakthroughs are expected to depend on AI as a foundational technology. AI is fast evolving with the potential to improve business operations, enhance public safety, and contribute to broader social progress.[2] At the same time, there will be issues, some foreseen, and many that will evolve alongside the technology itself. Regulating AI through traditional governance systems is challenging due to its pervasive and evolving nature. Instead, a degree of flexibility to promote innovation while ensuring security has often been provided through a variety of "soft-law," or non-binding, instruments.

1 Yanqing Duan, John S Edwards and Yogesh K Dwivedi, 'Artificial Intelligence for Decision Making in the Era of Big Data–Evolution, Challenges and Research Agenda' (2019) 48 International Journal of Information Management 63. doi:10.1016/j.ijinfomgt.2019.01.021.

2 Daniel Castro and Joshua New, *The Promise of Artificial Intelligence* (Center for Data Innovation 2016) 32-5.

Statutory law, which is the term used for written laws passed by a legislative body, is intended to give people and organisations a framework within which it is supposed to operate. Yet the regulation of AI's development and application has become increasingly complex due to society's growing reliance.[3]

Statutory law has long served as the cornerstone of legal systems, but in the era of artificial intelligence, these shortcomings have come to attention.[4] This research will examine the limitations of statutory law in the context of AI and explore how technological advances are disrupting established legal norms. It is becoming clear that statutory law alone cannot adequately handle the ethical, societal, and economic consequences of AI.

The widespread deployment of AI across diverse industries poses pressing issues of control and governance.[5] The rapid development of AI systems frequently surpasses the creation of related legal frameworks, posing significant challenges for legislators. Efforts to establish accountability and transparency are hindered by the dynamic nature of AI algorithms, which evolve and adapt within decision-making processes.[6] Regulatory efforts are further complicated by technical challenges, including bias and interpretability.

This research seeks to explore how statutory law can be adapted to effectively regulate artificial intelligence, striking a balance between fostering technological innovation and ensuring societal responsibility, while acknowledging its inherent limitations in addressing AI's unique ethical, social, and economic challenges.

2 METHODOLOGY AND RESEARCH APPROACH

This study examines the role of statutory law in AI regulation across several jurisdictions, employing a comparative legal analysis methodology in conjunction with doctrinal research methods. To explore how various legal systems address AI governance issues, the research employs a qualitative research methodology, combining a systematic examination of legislative texts, regulatory proposals, and court rulings.

The methodological framework consists of a few essential elements. First, doctrinal legal research forms the foundation of the analysis, involving a thorough examination of primary legal sources such as laws, rules, and case law pertaining to AI governance. This method

3 Laura F Edwards, *The People and their Peace: Legal Culture and the Transformation of Inequality in the Post-Revolutionary South* (University of North Carolina Press 2014).

4 Edward L Rubin, 'Law and Legislation in the Administrative State' (1989) 89(3) Columbia Law Review 369.

5 Lawrence B Solum, 'Artificially Intelligent Law' (2019) 1 BioLaw Journal 53. doi:10.15168/2284-4503-351.

6 Araz Taeihagh, 'Governance of Artificial Intelligence' (2021) 40(2) Policy and Society 137. doi:10.1080/14494035.2021.1928377.

reveals weaknesses in existing regulatory frameworks and permits a methodical analysis of how current legal frameworks respond to emerging AI challenges. Comparative legal analysis, focusing on the European Union's AI Act in conjunction with the latest frameworks in the US, UK, Brazil, and China, offers insights into various regulatory philosophies and approaches across jurisdictions.

The research methodology also incorporates analysis of secondary sources, including academic literature, policy documents, and industry reports, to provide contextual understanding of regulatory challenges and opportunities. This multi-source approach ensures comprehensive coverage of both theoretical frameworks and practical implementation challenges in AI regulation. The IRAC method (Issue, Rule, Application, and Conclusion) is used as an analytical framework to structure the legal analysis and ensure systematic examination of AI regulatory challenges.

Data collection was conducted through a systematic review of legislative texts, regulatory proposals, and academic commentary published between 2018 and 2024. This temporal scope was selected to encompass the most significant period of AI regulatory development globally, beginning with the European Union's initial AI regulatory proposals and extending to recent legislative initiatives in multiple jurisdictions.

3 HISTORICAL CONTEXT AND LEGISLATIVE EVOLUTION

Throughout history, the development of AI has been influenced by broader socio-political and economic environments, which in turn have shaped public attitudes and governmental responses. Early debates surrounding AI centred on how society would change as a result of it; however, concerns about job displacement and ethical dilemmas soon emerged, particularly in relation to productivity gains.

Changes in the availability of legal information have had an impact on the evolution of the legal services delivery system. With the invention of the CD-ROM in the 1990s, the use of digital resources and libraries gained momentum in the process of evolution.[7] Today, nearly all law firms rely on digital legal materials.[8] Predictive algorithms represent the most recent stage of this evolution, enabling lawyers to navigate complex legal dilemmas and identify and synthesise relevant information.

7 Ansgar Koene and others, *A Governance Framework for Algorithmic Accountability and Transparency* (EU 2019). doi:10.2861/59990.

8 George Stachokas, *The Role of the Electronic Resources Librarian* (Chandos 2019). doi:10.1016/C2018-0-02157-X.

Computer-assisted legal research, pioneered by Westlaw and Lexis in 1976, has become a cornerstone of legal practice. Most legal research is now conducted online,[9] replacing reliance on physical libraries. Digital access to legal texts allows attorneys to locate pertinent sources through keyword searches, significantly reducing the time once required to manually consult indices and read through each source individually.[10] While larger institutions rely heavily on proprietary data providers such as Westlaw, Lexis, or Bloomberg, others turn to publicly available sources.[11]

Traditionally, lawyers have been regarded as highly competent individuals trained to identify the relevant facts, frame pertinent legal issues, and predict the likely outcome of the case.[12] Attorneys apply judgment to evaluate the merits of a case and determine the best course of action by drawing on expertise and intuition. For a long time, such tasks were considered the exclusive domain of highly qualified specialists/ lawyers.

However, modern advances in AI have challenged long-held beliefs about human knowledge, particularly in the areas of machine learning and natural language processing. It is clear that the impact of data-driven analysis extends to the practice of law.[13] Litigation itself is gradually evolving, with disputes increasingly being resolved "in the shadow of the law", where settlement outcomes are shaped by the likely decision a court would reach. Predictions made by algorithms are repeatable by others.

The legislative development of AI can be traced through a series of significant events in several countries. Early regulations sought to advance R&D while maintaining ethical and safety standards. As AI applications spread across several industries, regulators faced new challenges concerning algorithmic transparency, cybersecurity, and data privacy.[14] In response, legislative solutions have been progressively updated to consider new dangers and public concerns. The definition of AI itself has remained fluid, shifting with technological advances.[15] In recent years, there has been a growing number of measures

9 Richard Susskind and Richard E Susskind, *Tomorrow's Lawyers: An Introduction to your Future* (OUP 2023).

10 Samuel Maireg Biresaw, 'The Impacts of Artificial Intelligence on Research in the Legal Profession' (2023) 5(1) International Journal of Law and Society 53. doi:10.11648/j.ijls.20220501.17.

11 F Allan Hanson, 'From Key Numbers to Keywords: How Automation Has Transformed the Law' (2002) 94 Law Library Journal 563.

12 Taryn Marks, 'John West and the Future of Legal Subscription Databases' (2015) 107(3) Law Library Journal 377. doi:10.2139/ssrn.2441734.

13 Cass R Sunstein, *Legal Reasoning and Political Conflict* (OUP 2018).

14 Melanie Mitchell, *Artificial Intelligence: A Guide for Thinking Humans* (Penguin UK 2019).

15 Peter Cihon, 'Standards for AI Governance: International Standards to Enable Global Coordination in AI Research & Development' (*Centre for the Governance of AI (GovAI)*, 17 April 2019) <https://www.governance.ai/research-paper/standards-for-ai-governance-international-standards-to-enable-global-coordination-in-ai-research-development> accessed 20 April 2025. ; Tarek Abo El-Wafa, 'The Jurisdiction of the UAE Federal Supreme Court on Constitutional Interpretation' (2021) 38(1–2) Arab Law Quarterly 192. doi:10.1163/15730255-bja10098.

to improve algorithmic accountability and transparency, including inquiries for regulatory monitoring of AI systems, particularly in high-stakes sectors such as criminal justice, finance, and healthcare.[16]

China's AI regulatory framework emphasises enhancing norms for scientific and technological ethics, focusing on data security protection systems and balancing fair competition with innovation encouragement. It addresses ethical concerns, attribution of liability, and the prevention of intellectual property monopolies. A central component is the *Interim Administrative Measures for Generative Artificial Intelligence Services,*[17] which plays a key role in ensuring governance mechanisms work together to build trust and accountability in the AGI industry.[18]

Nevertheless, China's current AI regulatory framework remains limited in scope, particularly in its lack of comprehensive definitions and protections for data rights. While it underscores the need for specialised intellectual property protection for data and the adoption of anti-monopoly measures to prevent misuse and monopolistic practices, it does not provide a fully developed legal framework. A more legal framework—one that recognises data as a form of intellectual property, acknowledges its dual public and proprietary nature, and strengthens digital IP protections alongside ethical guidelines for AI—is still needed.[19]

Brazil's approach is primarily outlined in Bill 21/2020,[20] which sets out the objectives and foundational principles for the development and use of AI, including ethical considerations. However, the bill does not explicitly regulate data privacy or intellectual property. Critics argue that its abstract provisions and limited number of articles risk overlooking essential aspects such as data protection and ethical guidelines, necessitating further discussion and refinement in the Senate.[21]

16 Sofia Samoili and others, *AI Watch: Defining Artificial Intelligence 2.0* (Publications Office of the EU 2021). doi:10.2760/01990.

17 Cybersecurity Administration of China and others, 'Interim Administrative Measures for Generative Artificial Intelligence Services' (13 July 2023) <https://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm> accessed 20 April 2025.

18 Niklas Kossow, Svea Windwehr and Matthew Jenkins, *Algorithmic Transparency and Accountability* (Transparency International 2021).

19 Bing Chen and Jiaying Chen, 'China's Legal Practices Concerning Challenges of Artificial General Intelligence' (2024) 13(5) Laws 60. doi:10.3390/laws13050060.

20 Brazil Bill 21/2020 'On a Legal Framework for Artificial Intelligence' (4 February 2020) <https://digitalpolicyalert.org/change/621-bill-2120-on-a-legal-framework-for-artificial-intelligence> accessed 20 April 2025.

21 Xiao Han and Nabeel Mahdi Althabhawi, 'Establishment of Data Intellectual Property Rights and Anti-Monopoly Regulation in China' (2024) 34(2) Jurnal Undang-Undang dan Masyarakat 190. doi:10.17576/juum-2024-3402-13.

While previous sections outlined individual jurisdictional approaches, a systematic comparison reveals fundamental divergences in how major economies balance innovation and responsibility through statutory frameworks:

The EU's AI Act (2024)[22] exemplifies a risk-based hierarchical model that prohibits unacceptable practices, such as social scoring and imposes stringent ex-ante requirements for high-risk systems, including conformity assessments and fundamental rights impact evaluations. Unlike the US's sectoral approach, the EU centralises enforcement through a European AI Office, creating uniform compliance burdens. Critics argue this may stifle startups lacking resources for compliance,[23] while proponents highlight its strong emphasis on safeguarding fundamental rights.

By contrast, the US regulates AI through fragmented sectoral agencies such as the FTC, FDA, and NTSB, relying largely on non-binding frameworks like the NIST AI RMF and the Blueprint for an AI Bill of Rights. Proposed legislation, such as the Algorithmic Accountability Act (2023)[24], focuses narrowly on impact assessments in specific contexts such as hiring and housing. This avoids EU-style centralised burdens but creates regulatory uncertainty and enforcement gaps. Further, state-level initiatives—such as California's AB 331—push stricter rules, risking a fragmented "patchwork" regime. The absence of federal AI liability laws leaves accountability reliant on tort law, creating ambiguity for autonomous systems.

China's approach merges aggressive state investment in AI R&D with strict control mechanisms, exemplified by the Generative AI Interim Measures (2023). Its "negative list" system prohibits challenges to state authority while promoting industrial dominance in non-sensitive sectors. Unlike Western models that prioritise individual rights, China emphasises data sovereignty, social stability, and alignment with "socialist core values." This facilitates rapid scaling of state-approved innovations but restricts algorithmic transparency and independent oversight.

Brazil's Bill 21/2020 mirrors EU principles (human oversight, non-discrimination) but lacks implementation mechanisms. The bill coexists with the LGPD (GDPR-inspired data law), creating potential conflicts between data minimisation and AI training needs. More broadly,

22 Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance) [2024] OJ L 1689/1 <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689> accessed 20 April 2025.

23 Michael Veale and Frederik Zuiderveen Borgesius, 'Demystifying the Draft EU Artificial Intelligence Act' (2021) 22(4) Computer Law Review International 97.

24 US 2892 Algorithmic Accountability Act (21 September 2023) <https://www.congress.gov/bill/118th-congress/senate-bill/2892> accessed 20 April 2025.

emerging economies often adopt EU-style principles but encounter capacity gaps in enforcement. In several cases, such as India's Digital India Act, governments prioritise "sovereign AI" infrastructure over ethics frameworks, reflecting divergent innovation-responsibility weightings.

4 ROLE OF STATUTORY LAW IN REGULATING ARTIFICIAL INTELLIGENCE

International cooperation in AI governance increasingly centres on initiatives to establish technological and ethical standards that ensure the responsible development and application of AI systems.[25] Organisations, including the International Telecommunication Union, the International Organisation for Standardisation, and the Organisation for Economic Co-operation and Development, have undertaken initiatives to create AI guidelines, standards, and principles that support accountability, openness, equity, and user-centred design.

Within this global landscape, statutory law outlines rules for the creation, implementation, and use of AI systems. It provides a legal framework within which individuals and organisations operate and serves several key regulatory functions.[26] First, clear guidelines and norms for the creation and use of AI systems are one of the key functions of statutory law in regulating the emerging technologies.[27] To ensure that AI systems do not endanger the public's health or safety, statutory law can set minimal safety criteria.[28] To further ensure that personal information and civil liberties are recognised, it can also establish rules for the gathering and use of data in AI systems.[29]

The function of statutory law includes regulating AI-systems by emphasising their liability and accountability.[30] If an AI system affects someone, the creator, the user, or the AI system

25 Mayara Rayssa da Silva Rolim, Daniella Maria dos Santos Dias and Gabriel Napoleão Velloso Filho, 'Regulation of Algorithms in Artificial Intelligence Systems: A Possible Proposal for Brazil?' (2024) 17(2) Contribuciones a Las Ciencias Sociales e4924. doi:10.55905/revconv.17n.2-006.

26 Cihon (n 17).

27 Benjamin Myles Cheatham, Kia Javanmardian and Hamid Samandari, 'Confronting the Risks of Artificial Intelligence' (2019) 2 McKinsey Quarterly 38.

28 Alan FT Winfield and Marina Jirotka, 'Ethical Governance is Essential to Building Trust in Robotics and Artificial Intelligence Systems' (2018) 376(2133) Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 20180085. doi:10.1098/rsta.2018.0085.

29 Andrea Romaoli Garcia, 'AI, IoT, 'Big Data, and Technologies in Digital Economy with Blockchain at Sustainable Work Satisfaction to Smart Mankind: Access to 6th Dimension of Human Rights' in Nuno Vasco Moreira Lopes (ed), *Smart Governance for Cities: Perspectives and Experiences* (Springer 2020) 83. doi:10.1007/978-3-030-22070-9_6.

30 Alessandro Mantelero and Maria Samantha Esposito, 'An Evidence-Based Methodology for Human Rights Impact Assessment (HRIA) in the Development of AI Data-Intensive Systems' (2021) 41 Computer Law & Security Review 105561. doi:10.1016/j.clsr.2021.105561. Tarek Abo El-Wafa, Ahmed Khalil and Adham Hashish, 'Parliamentary question: Insights from the Federal National Council in the UAE' (2024) 10(6) *Heliyon* e27671. doi:10.1016/j.heliyon.2024.e27671.

itself could all be held accountable according to statutory law.[31] This is particularly crucial when AI systems are applied in high-risk industries, such as healthcare or transportation.[32]

Statutory law also contributes to limiting the impact of AI on employment and the economy. To prevent discrimination, statutory legislation should explore rules for the application of AI to hiring and promotion decision-making processes. Additionally, it can establish rules for the application of AI in fields where job displacement is a concern.

4.1. Statutory Laws and Their Limitations

The limitations of statutory law in addressing challenges posed by AI have come to light. As AI systems grow in complexity and autonomy, their regulation becomes more difficult.[33] It might equally be challenging for regulators to comprehend how they operate or recognise possible concerns.[34] Considering how quickly AI technology is developing, it is challenging for statutory law to keep up with and adjust to new advances.[35]

In principle, rules should be relied upon to regulate human conduct, as the law must strike a balance between flexibility, certainty, and reliability. Compared to principles, rules are more definite and easier to apply consistently. Law remains one of society's most crucial instruments for shaping behaviour, offering rewards for certain actions and penalties for others, influencing the creation of social institutions.

However, the law can also distort individual decision-making. While awareness of legal consequences can be helpful, it may constrain a person's ability to act in a way that reflects their genuine preferences, moral convictions, and economic interests. An overemphasis on legal compliance risks fostering excessive strategic thinking and manipulative conduct. Ultimately, this is damaging to psychological well-being, distributive fairness, autonomy, and efficiency. Achieving an optimal equilibrium between the law's beneficial function and its potential to distort behaviour is a challenging matter.

Insolvency law provides an example where concealment of legal rules is sometimes justified due to ex-ante strains.[36] When taking out a loan and investing its proceeds,

31 Bernd W Wirtz, Jan C Weyerer and Benjamin J Sturm, 'The Dark Sides of Artificial Intelligence: An Integrated AI Governance Framework for Public Administration' (2020) 43(9) International Journal of Public Administration 818. doi:10.1080/01900692.2020.1749851.

32 Shlomit Yanisky-Ravid, 'Generating Rembrandt: Artificial Intelligence, Copyright, and Accountability in the 3A Era: The Human-like Authors Are Already Here: A New Model' [2017] Michigan State Law Review 659. doi:10.2139/ssrn.2957722.

33 Amy Rankin and others, 'Resilience in Everyday Operations: A Framework for Analyzing Adaptations in high-Risk Work' (2014) 8(1) Journal of Cognitive Engineering and Decision Making 78. doi:10.1177/1555343413498753..

34 Duan, Edwards and Dwivedi (n 1).

35 Matthew U Scherer, 'Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies' (2016) 29(2) Harvard Journal of Law & Technology 353. doi:10.2139/ssrn.2609777.

36 Richard M Re and Alicia Solow-Niederman, 'Developing Artificially Intelligent Justice' (2019) 22 Stanford Technology Law Review 242.

debtors bear full responsibility for repayment obligations.[37] However, if unable to repay, the law offers relief through various legal solutions to the creditor. If debtors were fully aware of these remedies beforehand, their behaviour might become more opportunistic or uncertain. In such cases, shielding the specifics of ex post legal remedies can be justified to preserve responsibility and fairness.

There can never be a definitive settlement of the relative jurisdictions of legislators and courts. Like other aspects of political organisation, this relationship is open to ongoing interpretation and discussion. In substantive criminal law, three principles govern the relationship between legislatures and courts: the legality principle, or nulla poena sine lege;[38] the constitutional theory of void-for-vagueness;[39] and the principle of strict construction, which dictates that courts must always construe criminal statutes in a way that favours the accused when there is any remaining ambiguity. Taken as a whole, these principles reflect a cautious approach to judicial lawmaking, limiting the judiciary's role in the creation of criminal law.

At the same time, AI underscores the importance of aligning legal systems with social considerations and public policy goals.[40] The societal influence of AI is intricate and calls for a wide range of public policy solutions, from employment prospects to ethical issues. One major concern is the disruption of labour markets, which has sparked conversations about workforce transition plans, universal basic income, and retraining initiatives. Equity and fairness are often compromised by AI-driven algorithms, leading to issues of prejudice, discrimination, and unequal access to opportunities.

The formulation of laws regulating AI is complicated by the inherent restrictions imposed by statutory law.[41] While statutory law is intended to give people and organisations a framework within which to operate,[42] it may not always be able to foresee the particular difficulties brought on by innovative and quickly developing AI technology. Therefore, legal loopholes frequently emerge, undermining the effectiveness of statutory regulation and necessitating supplementary reforms.[43]

37 Sandeep Gopalan and Michael Guihot, 'Recognition and Enforcement in Cross-Border Insolvency Law: A Proposal for Judicial Gap-Filling' (2015) 48 Vanderbilt Law Review 1225.

38 George G Triantis, 'Theory of the Regulation of Debtor-in-Possession Financing, A' (1993) 46 Vanderbilt Law Review 901.

39 Paul H Robinson, 'Fair Notice and Fair Adjudication: Two Kinds of Legality' (2005) 154 University of Pennsylvania Law Review 335.

40 Peter L Strauss, 'Legislative Theory and the Rule of Law: Some Comments on Rubin' (1989) 89 Columbia Law Review 427.

41 Mihail C Roco and William S Bainbridge, 'The New World of Discovery, Invention, and Innovation: Convergence of Knowledge, Technology, and Society' (2013) 15 Journal of Nanoparticle Research 1946. doi:10.1007/s11051-013-1946-1.

42 Scherer (n 35).

43 Ruth Suseela Meinzen-Dick and Rajendra Pradhan, *Legal Pluralism and Dynamic Property Rights* (CAPRi Working Paper no 22, International Food Policy Research Institute 2002).

The rise of AI presents a variety of challenges to conventional legal systems.[44] AI is increasingly applied in decision-making processes such as personnel selection, credit assessment, and aspects of legal analysis. However, these applications raise serious concerns about accountability, transparency, and fairness.[45] More innovative applications—such as driverless cars, AI-powered medical diagnostics, and autonomous drones—introduce regulatory questions that conventional legislation might not be able to handle as AI technologies advance and diversify. For example, the deployment of autonomous vehicles highlights unresolved issues regarding responsibility, safety requirements, and regulatory oversight, since current transportation laws do not adequately account for their unique capabilities and difficulties.

The broad application of AI in a variety of industries, such as healthcare, banking, and criminal justice, highlights the necessity of industry-specific laws designed to handle risks and issues unique to each industry.[46] However, the fragmented nature of regulatory initiatives, combined with the rapid pace of technological advancement, can lead to gaps and inconsistencies that undermine effective enforcement. Moreover, the global nature of AI development and application complicates governance:[47] diverging national frameworks and standards risk creating obstacles to innovation and interoperability. This makes international cooperation and harmonising regulatory frameworks essential for closing legal loopholes and ensuring accountability, safety and fairness in the use of AI technologies.

A fundamental limitation of statutory law is its reactive character. Statutory laws typically emerge in response to an existing issue or problem, which makes them resistant to adapting proactively to new developments. AI presents particular challenges in this regard: as AI evolves rapidly, new ethical and legal concerns continually emerge, often faster than legislatures can respond. Consequently, statutory law risks lagging behind technological change, leaving legal systems unprepared to address the unique issues brought on by the usage of AI.

The disadvantage of statutes is that they are often overly restrictive. Although statutory law is intended to be broadly applicable across a range of scenarios, this generality can make it difficult to address specific situations. This is particularly problematic for AI, which is routinely used in varied situations that do not neatly align with established legal frameworks.

44 Miriam C Buiten, 'Towards Intelligent Regulation of Artificial Intelligenceبمعه' (2019) 10(1) European Journal of Risk Regulation 41.

45 Paulius Čerka, Jurgita Grigienė and Gintarė Sirbikytė, 'Liability for Damages Caused by Artificial Intelligence' (2015) 31(3) Computer Law & Security Review 376. doi:10.1016/j.clsr.2015.03.008.

46 Corinne Cath, 'Governing Artificial Intelligence: Ethical, Legal and Technical Opportunities and Challenges' (2018) 376(2133) Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 20180080. doi:10.1098/rsta.2018.0080.

47 Shalini Rai, 'Legal Liability Issues and Regulation of Artificial Intelligence' (thesis, 2022).

Statutory law frequently faces limitations arising from its dependence on human interpretation. Because laws are expressed in natural language, they are inherently open to multiple interpretations.[48] This limitation is particularly salient in the context of AI, where algorithmic decisions can be difficult to understand or explain, potentially producing inconsistent and ambiguous legal outcomes. In *State v. Loomis*, the Wisconsin Supreme Court acknowledged the defendant's challenge to the use of the COMPAS risk assessment algorithm in sentencing. While upholding its use, the court mandated specific warnings for judges, highlighting judicial recognition of statutory law's struggle with algorithmic opacity and potential bias.[49] It may be challenging to prove that these biases exist, though, given the often inaccessible nature of AI decision-making processes.

The rigidity and constrained nature of statutory law are key factors limiting its ability to address the ethical and moral implications of AI.[50] Legal systems often struggle to keep pace with the remarkable rate at which AI technologies are developing,[51] and the inflexible terminology of statutes may not take unforeseeable events or the broader ethical ramifications of AI into account.[52]

Statutory law also relies heavily on human interpretation, which can be perplexing or unpredictable in situations involving intricate moral and ethical dilemmas.[53] Different interpreters may reach divergent conclusions, resulting in inconsistent or contradictory applications of the law.[54] Furthermore, statutory law typically concentrates on addressing certain problems, whereas AI has the potential to impact multiple facets of society.[55] This means that the profound ethical and moral consequences of AI may not be fully addressed.

The regulation of AI through statutory legislation faces a number of restrictions. Some of these restrictions include:

1) **Lack of clarity**: Terms associated with AI, such as machine learning algorithms, neural networks, and deep learning, often lack precise legal definitions, making it challenging to apply existing statutes to developing technology.

48 Taeihagh (n 6).

49 Mark Greenberg, 'Legal Interpretation and Natural Law' (2020) 89(1) Fordham Law Review 109.

50 *State v Loomis* 881 NW2d 749 [2016] Supreme Court of Wisconsin.

51 Olivia J Erdélyi and Judy Goldsmith, 'Regulating Artificial Intelligence: Proposal for a Global Solution' (AIES '18: AAAI/ACM Conference on AI, Ethics, and Society, New Orleans LA USA, 2-3, February 2018) 95.

52 Yogesh K Dwivedi and others, 'Artificial Intelligence (AI): Multidisciplinary Perspectives on Emerging Challenges, Opportunities, and Agenda for Research, Practice and Policy' (2021) 57 International Journal of Information Management 101994. doi:10.1016/j.ijinfomgt.2019.08.002.

53 Margarita Robles Carrillo, 'Artificial Intelligence: From Ethics to Law' (2020) 44(6) Telecommunications Policy 101937. doi:10.1016/j.telpol.2020.101937.

54 Scherer (n 35).

55 Ken Kress, 'Legal indeterminacy' (1989) 77(2) California Law Review 283. doi:10.15779/Z380B17.

2) **Slow legislative process:** Enacting or amending statutes is typically a lengthy process. Consequently, statutory law often lags behind the rapid development of AI, creating regulatory gaps and loopholes.

3) **Limited application:** Statutory law often focuses on particular regulatory areas, such as liability, security, or privacy. This narrow focus may not be sufficient to address the complex and numerous problems that arise from the creation and application of AI.

4) **Inflexibility**: Once enacted, statutes can be difficult to amend or modify. This rigidity hinders timely responses to new and developing AI-related concerns, and the variation between national legal systems can complicate efforts to regulate AI on a global level.

5) **Enforcement challenges**: Enforcing statutory law can be challenging, particularly in the age of rapidly developing technologies like artificial intelligence. The complexity of AI systems, combined with ambiguities in existing legal provisions, can make it challenging to identify and hold violators accountable.

4.2. The Difficulties of Juggling Innovation and Responsibility

One of the central challenges in balancing innovation and accountability in AI governance is promoting the development of AI technologies while ensuring their use is responsible and ethical. [56]This entails addressing the statutory law's restrictions on the regulation of AI, the quick advancement of technology, and the broader moral, societal and economic implications of AI.

Three foundational arguments underlie key legal doctrines, including the vagueness theory, the rule of rigid construction, and nulla poena sine lege.[57] First, judicial innovation is rendered illegitimate when popular sovereignty is linked to legislative supremacy, reflecting the principle of "separation of powers" in modern constitutionalism. Second, it is unfair to penalise behaviour that was not previously classified as criminal, emphasising "notice" and "fair warning". Third, concerns about biased or arbitrary application of the criminal code highlight the importance of legal formalism in constraining unchecked discretion.

The so-called "rule of law" underlies both the vagueness theory and *nulla poena sine lege*. Yet there is danger in invoking this term, as it has become a highly malleable political catchphrase. Too often, the rule of law is conflated with the rule of good law, turning it into a blanket assertion of virtue within a legal framework. Legal theorists and philosophers,

56 Abdulfattah Yaghi, Tarek Abo El-Wafa and Ali A Al Ahbabi, 'Exploration of principal–agent theory in a consultative policy-making context' (2025) 12 *Humanities and Social Sciences Communications* 1419. doi:10.1057/s41599-025-05648-4.

57 Bruce G Buchanan and Thomas E Headrick, 'Some Speculation about Artificial Intelligence and Legal Reasoning' (1970) 23(1) Stanford Law Review 40. doi:10.2307/1227753.

however, have developed the idea in unduly complex ways. At its core, the prohibition of arbitrary behaviour in the use of state power is symbolised by the rule of law. In the context of criminal law, it requires that agents of official coercion, to the extent possible, act in accordance with established rules—that is, publicly recognised, reasonably stable, and broadly applicable declarations of prohibited behaviour by the state.

The economic impact of AI is another critical factor shaping industrial and societal development. AI integration can enhance productivity and foster innovation, but it also raises challenges pertaining to competitiveness, market dynamics, and regulatory control. Automation has the potential to transform the labour market, triggering issues such as job displacement, the need for skill retraining, and income inequality. At the same time, AI raises significant ethical and societal concerns, particularly when algorithms are used to make judgments in hiring, credit or criminal justice. Bias in algorithmic decision-making can have profound effects on people and communities.

The key considerations for AI governance are described below:

4.2.1. Privacy & Social Implications

Privacy is a critical area of concern that deserves attention while developing AI.[58] As machine learning algorithms evolve, they have the capacity to collect, analyse, and store vast amounts of personal data, raising questions about how such data is used and safeguarded.[59] This issue is particularly acute in healthcare, where AI is being applied to create innovative treatments and actions, creating concerns about the security and privacy of patient data.[60]

The ethical governance of AI is further complicated by the limitations of statutory law. While statutory frameworks provide a structured environment in which people and organisations can function, they may be inadequate to address the unique ethical challenges posed by AI.[61] Emerging AI systems can generate new types of harm that aren't protected by current legal frameworks, or they could cause concerns about the accountability and duty of those who utilise AI.[62] The ability of AI technologies to make

58 John Calvin Jeffries Jr, 'Legality, Vagueness, and the Construction of Penal Statutes' (1985) 71(2) Virginia Law Review 189. doi:10.2307/1073017.

59 Yi Zhang and others, 'Ethics and Privacy of Artificial Intelligence: Understandings from Bibliometrics' (2021) 222(24) Knowledge-Based Systems 106994. doi:10.1016/j.knosys.2021.106994.

60 Michael I Jordan and Tom M Mitchell, 'Machine Learning: Trends, Perspectives, and Prospects' (2015) 349(6245) Science 255. doi:10.1126/science.aaa841.

61 Richard J Chen and others, 'Synthetic Data in Machine Learning for Medicine and Healthcare' (2021) 5(6) Nature Biomedical Engineering 493. doi:10.1038/s41551-021-00751-8.

62 Cass R Sunstein, 'On the Expressive Function of Law' (1996) 144(5) University of Pennsylvania Law Review 2021. doi:10.2307/3312647.

choices and forecasts without human input introduces complex ethical dilemmas,[63] particularly related to privacy, bias, and discrimination.[64]

AI systems often process extensive personal datasets, including biometric data, location data, and browser history, to generate predictions and judgments about individuals.[65] Without people's awareness or consent, this data can be gathered and utilised, raising privacy, security and monitoring concerns.

4.2.2. Bias and Discrimination Concerns

Machine learning algorithms can be trained on biased datasets, which may result in biased projections and conclusions, raising issues about prejudice and discrimination.[66] Such biases can disproportionately affect particular groups, including minorities or people with disabilities. Judicial awareness of the tension between technological tools and fairness is evident in cases such as *State v. Loomis* (Wisconsin) and *Commonwealth v. Scantling* (Massachusetts), where courts grappled with the admissibility and fairness of algorithmic risk assessment tools in criminal sentencing, directly confronting the bias and responsibility challenges inherent in AI adoption.[67]

Effective AI governance necessitates the implementation of risk evaluation and compliance applications, enabling businesses to recognise, reduce, and manage risks posed by AI technologies while maintaining compliance with legal and ethical obligations. Risk assessment involves systematically evaluating possible hazards and weaknesses throughout the AI lifecycle—from data collection and model development to deployment and operation. This process locates sources of potential harm, including algorithmic bias, security flaws, data privacy violations, and unintended consequences, and evaluates their likelihood and potential impacts on relevant parties.

Another concern is the potential for AI systems to cause harm to people or society.[68] For instance, AI-powered medical diagnosis systems may generate false diagnoses, which

63 Nithesh Naik and others, 'Legal and Ethical Consideration in Artificial Intelligence in Healthcare: Who Takes Responsibility?' (2022) 9 Frontiers in Surgery 266. doi:10.3389/fsurg.2022.862322.

64 Thomas Davenport and others, 'How Artificial Intelligence Will Change the Future of Marketing' (2020) 48 Journal of the Academy of Marketing Science 24. doi:10.1007/s11747-019-00696-0.

65 Syeda Faiza Nasim, Muhammad Rizwan Ali and Umme Kulsoom, 'Artificial Intelligence Incidents & Ethics a Narrative Review' (2022) 2(2) International Journal of Technology, Innovation and Management 52. doi:10.54489/ijtim.v2i2.80.

66 Jeroen Van den Hoven and others, 'Privacy and Information Technology', *Stanford Encyclopedia of Philosophy* (2014) <https://plato.stanford.edu/entries/it-privacy/> accessed 20 April 2025.

67 Jessica K Paulus and David M Kent, 'Predictably Unequal: Understanding and Addressing Concerns that Algorithmic Clinical Prediction May Increase Health Disparities' (2020) 3(1) NPJ Digital Medicine 99. doi:10.1038/s41746-020-0304-9.

68 *State v Loomis* (n 51); *Commonwealth v Scantling* 24 NE 3d 1064 [2015] Supreme Judicial Court of Massachusetts.

could potentially harm patients. Similarly, autonomous weapons systems by AI have the potential to make fatal judgments without human intervention, raising questions regarding ethics and morality.[69]

The intersection of human rights and AI regulation has become increasingly critical, as AI technologies pose significant challenges to fundamental human rights like privacy, freedom of speech, equality before the law, and access to justice. Ensuring that AI development and deployment conform to the basic principles of human rights, particularly as AI becomes more integral to the criminal justice system, healthcare system, workplaces, and social services.

Concerns have arisen regarding the potential for AI systems to reinforce or amplify biases and inequities at the junction of human rights. The tension between AI deployment and fundamental rights is starkly illustrated in pending cases before international bodies. For instance, in *AlgorithmWatch Schweiz and others v. Switzerland* before the European Court of Human Rights, the plaintiffs challenge the lack of human review and transparency in fully automated systems used for significant public decisions. This case tests the boundaries of Article 8 (Privacy) and Article 6 (Fair Trial) of the European Convention on Human Rights.[70] AI algorithms trained on biased data may generate discriminatory outcomes, potentially resulting in unfair treatment and infringements on people's rights to equal protection under the law.

4.2.3. Economic Repercussions

Technology based on AI can drastically change a number of industries and open up new business opportunities,[71] but it also prompts questions regarding the economic effects of its creation and application, notably with regard to the loss of jobs and the concentration of wealth.

AI can automate many operations currently performed by humans, leading to significant job displacement,[72] especially in industries that involve repetitive tasks, such as manufacturing or data entry. Businesses that successfully adapt and use AI technology may experience an increase in productivity and profitability, potentially increasing wealth and market domination.[73] However, this may exacerbate economic inequality, concentrating wealth among a small number of businesses and individuals.

69 Andreas Kaplan and Michael Haenlein, 'Rulers of the World, Unite! The Challenges and Opportunities of Artificial Intelligence' (2020) 63(1) Business Horizons 37. doi:10.1016/j.bushor.2019.09.003.

70 Sabriya Alam and others, 'Unmanned and Autonomous Weapons Systems: Practices and Related Policy' (2020) 2(1) PPRI Student Papers in Public Policy 7.

71 *AlgorithmWatch Schweiz and others v Switzerland* App no 52652/21 (ECtHR, 24 January 2023).

72 Dwivedi (n 52).

73 Thomas H Davenport and Rajeev Ronanki, 'Artificial Intelligence for the Real World' (2018) 96(1) Harvard Business Review 108.

The economic impact of AI extends beyond particular businesses or industries. Widespread use of AI may have macroeconomic implications, including shifts in the labour market, changes in supply and demand, and variations in rates of economic expansion.[74]

AI also challenges the traditional legal framework by generating new kinds of data and information.[75] Existing statutory laws often struggle to anticipate emerging issues in areas such as intellectual property, privacy, and data protection, which may therefore be overlooked and inadequately addressed.[76]

Furthermore, AI has the potential to cast doubt on accountability under the law.[77] For example, it might not be apparent who is legally accountable for the harm caused if a medical diagnosis system powered by artificial intelligence makes a wrong diagnosis,[78] creating ambiguity and misunderstanding that statutory law may not be equipped to resolve.[79] High-profile cases, such as *Waymo LLC v. Uber Technologies, Inc.,* though settled, underscore the complex liability and intellectual property challenges arising from rapid AI development, particularly in self-driving car systems, and highlight the difficulty statutory frameworks face in definitively assigning responsibility.[80]

The fast pace and increasing complexity of AI development make it challenging for policymakers and legal professionals to enact laws capable of effectively governing AI creation and deployment.[81]

74 Spyros Makridakis, 'The Forthcoming Artificial Intelligence (AI) Revolution: Its Impact on Society and Firms' (2017) 90 Futures 46. doi:10.1016/j.futures.2017.03.006.

75 Richard B Freeman, 'Labour Market Institutions Without Blinders: The Debate Over Flexibility and Labour Market Performance' (2005) 19(2) International Economic Journal 129. doi:10.1080/10168730500080675.

76 Dirk Helbing, 'Societal, Economic, Ethical and Legal Challenges of the Digital Revolution: From Big Data to Deep Learning, Artificial Intelligence, and Manipulative Technologies' in Dirk Helbing (ed), *Towards Digital Enlightenment* (Springer 2019) 47. doi:10.1007/978-3-319-90869-4_6.

77 William M Landes and Richard A Posner, *The Economic Structure of Intellectual Property Law* (Harvard UP 2003).

78 Marten Risius and Kai Spohrer, 'A Blockchain Research Framework: What We (don't) Know, Where We Go from Here, and How We Will Get There' (2017) 59 Business & information systems engineering 385.

79 Stacy M Carter and others, 'The Ethical, Legal and Social Implications of Using Artificial Intelligence Systems in Breast Cancer Care' (2020) 49 The Breast 25. doi:10.1016/j.breast.2019.10.001.

80 Sara Gerke, Timo Minssen and Glenn Cohen, 'Ethical and Legal Challenges of Artificial Intelligence-Driven Healthcare' in Adam Bohr and Kaveh Memarzadeh (eds), *Artificial Intelligence in Healthcare* (Academic Press 2020) 295. doi:10.1016/B978-0-12-818438-7.

81 Case No 17-CV-00939-WHA *Waymo LLC v Uber Technologies, Inc*, Settlement Order (ND Cal, 7 February 2018).

5 LIMITATIONS OF LEGAL POLICIES IN REGULATING DEVELOPMENTS OF AI

Two considerations must be mentioned at the outset. First, there are different degrees of conformity to the law. It is difficult to envisage a legal system composed solely of precise, mechanical principles—and it would likely be undesirable even if possible. Some degree of discretion will always remain in the legal system.

Second, as AI is increasingly used to support or replace human decision-making, the question arises as to what kind of process should be afforded to individuals affected by such judgments. The growing prevalence of black-box machine-learning algorithms renders many machine decision-making virtually unintelligible. This opacity is compounded by the phenomenon of "automation bias," whereby individuals exhibit overconfidence in the judgments made by machines and display prejudice against challenges to those determinations. Although AI is often promoted for its potential to reduce costs and increase efficiency, these benefits remain uncertain in the face of obstacles, particularly if they include significant procedural rights, such as transparency and due process, that are poorly safeguarded.

Although there have been attempts to manage and regulate the development of AI, there are still some legal constraints that need to be taken into consideration. These restrictions include, among others:

1) **Inability to keep pace with technological change**: AI technology evolves rapidly, whereas lawmaking is inherently slow. New laws take time to draft and put into effect, and existing laws can swiftly become obsolete.

2) **Lack of global cooperation:** The growth and implementation of AI transcend national boundaries. In the absence of international coordination, fragmented legal frameworks risk creating inconsistencies, enforcement gaps, and jurisdictional challenges. These difficulties are exacerbated when infringing parties are positioned in another country or when the legal system is precarious.

3) **Ethical issues:** The possibility of bias, discrimination, and privacy invasion are a few of the serious ethical issues raised by AI.[82] These challenging ethical issues demand constant attention and assessment, and legal systems and regulations may find it difficult to handle them.

4) **Limited knowledge of AI**: Many legal professionals lack the technical know-how necessary to comprehend AI and its implications.[83] Their capacity to create and put into practice efficient legal regulations and policies may be constrained as a result.

82 Sharona Hoffman and Andy Podgurski, 'Artificial Intelligence and Discrimination in Health Care' (2019) 19(3) Yale Journal of Health Policy, Law, and Ethics 1.

83 Waleed Ali and Mohamed Hassoun, 'Artificial Intelligence and Automated Journalism: Contemporary Challenges and New Opportunities' (2019) 5(1) International Journal of Media, Journalism and Mass Communications 40. doi:10.20431/2454-9479.0501004.

6 INTERNATIONAL LEGAL DIFFICULTIES ARISING FROM THE USE OF AI

AI adoption generates a range of cross-border legal challenges that require careful attention.[84] In the United States, regulatory efforts regarding system-wide risk reduction in algorithmic decision-making have largely overlooked individual due process. There has been some agreement among recent legislative proposals from the United States; however, they have emphasised the need for systemic solutions rather than individual rights to contest, such as algorithmic impact studies or audits.

In contrast, regulators in Europe are approaching algorithmic decision-making from a comprehensive standpoint. The European Union's General Data Protection Regulation (GDPR),[85] which came into force in May 2018, provides certain individual rights for data subjects as well as systemic governance measures. Individuals whose decisions are subject to automated decision-making also have the ability to argue against specific findings. These rights also include access, transparency, amongst others. Similarly, the Council of Europe's updated data protection treaty specifies the right to appeal. The Council of Europe is an international organisation devoted to human rights, consisting of the member states of the European Union and a few non-EU nations.

Beyond Europe, the right to challenge AI outcomes is also gaining ground. The Organisation for Economic Co-operation and Development (OECD), an intergovernmental body influential in shaping global data protection regulations through its recommendations, is expected to extend its influence into AI governance as well. The "right to request a review of decisions taken" by AI is included in Brazil's extensive data protection law, enacted in 2018. Similarly, in November 2020, the Canadian Office of the Privacy Commissioner recommended amending Canada's data privacy law to introduce a right to challenge AI conclusions.

Despite these developments, several significant legal challenges arise from the use of AI:

1) **Intellectual property**: AI systems may be protected under trade secrets, copyrights, or other types of intellectual property protection.[86] When using another person's intellectual property, developers must take care to respect their rights and secure the appropriate licenses or permissions.

84 Erik Brynjolfsson and Andrew McAfee, 'Artificial Intelligence, for Real' (2017) 1 Harvard Business Review 1.

85 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L 119/1<http://data.europa.eu/eli/reg/2016/679/oj> accessed 20 April 2025.

86 Jennifer Cobbe and Jatinder Singh, 'Artificial Intelligence as a Service: Legal Responsibilities, Liabilities, and Policy Challenges' (2021) 42 Computer Law & Security Review 105573. doi:10.2139/ssrn.3824736.

2) **Liability**: AI raises complex questions of liability, especially in cases of accidents or errors brought on by algorithmic flaws, biases, or mismanagement. Developers and users must determine who should be held liable for any damage the systems may cause and ensure that sufficient insurance is in place to cover any potential risks. Traditional legal frameworks often struggle to sufficiently handle these unexpected difficulties, prompting calls for judicial precedents and clearer accountability standards.

3) **Jurisdiction**: As AI systems frequently operate across borders, they can lead to complicated jurisdictional difficulties. Developers and users must make sure they abide by the rules and laws of all pertinent jurisdictions.

4) **Prejudice**: AI systems trained on biased datasets risk generating discriminatory outcomes.

5) **Trade restrictions**: Where AI technologies have potential military or national security uses,[87] they may be subject to export controls and trade restrictions. Developers and users must therefore comply with all applicable export regulations.

7 CONCLUSIONS AND SUGGESTIONS

The role of statutory law in regulating AI is fundamental in striking a delicate balance between fostering innovation and ensuring responsibility. As AI continues to evolve at a rapid pace, the establishment of robust legal frameworks becomes essential to address the ethical, social, and economic implications that accompany its development. Moreover, effective regulation must be dynamic and adaptive, reflecting the evolving nature of technology while maintaining core principles of transparency, privacy, and security.

Notice and transparency obligations under the General Data Protection Regulation (GDPR) for AI have garnered more attention, especially the so-called "right to explanation," which has ignited an upsurge of scholarly discourse. Though the GDPR explicitly establishes the right to dispute, regulators have not yet offered substantial guidance on the nature of the right or how it should be exercised.

The essence of democracy is under threat from the growing use of AI in decision-making. It is crucial to use design approaches that incorporate judicial review concepts as a fundamental component of AI-driven architecture in order to restore human confidence in AI. However, AI cannot wholly replace human bias and is therefore not always accurate; instead, it may obscure bias behind layers of purportedly impartial mathematical authority. Algorithmic outcomes can be biased even when programmers do not intend to discriminate, and these problems manifest across diverse technologies. For example,

87 Frank A DeCosta III, 'Intellectual Property Protection for Artificial Intelligence' [2017] Westlaw Journal Intellectual Property <https://www.finnegan.com/en/insights/articles/intellectual-property-protection-for-artificial-intelligence.html> accessed 20 April 2025.

actuarial algorithms used in criminal sentencing— despite their simplicity—have been shown to perpetuate bias, discrimination, and inaccuracy.

AI decision-making further raises questions about "what it means to be human." By excluding human judgment, empathy, and contextual reasoning, both public and private institutions risk reducing individuals to numerical values. It is arguable that the dignity of the human subject of the judgment is compromised when human decision-makers are replaced by machines. Yet, it would be a mistake to assume that algorithms are inherently more flawed than human decision makers; judges, too, may act with prejudice or inaccuracy. The dignity of a human subject can also be harmed by discrimination by a human decision-maker.

However, the transition from human to AI or hybrid human-AI decision-making systems fundamentally changes the policy environment and its underlying values. For instance, AI decision-making transfers some policy decisions early on to algorithm designers, rather than allowing a human decision-maker to assess a specific individual's unique circumstances ex post. In many cases, policy choices remain embedded in the "black box" of the algorithm, potentially opaque even to its creators. Decision-making processes and who decides what vary. The lack of transparency undermines accountability and the outcomes of the decisions.

The transition from individual customisation to decisions based on categories may also accompany a shift in AI decision-making. This gives rise to an issue known as the "long-tail problem," when an AI incorrectly classifies "weird stuff that is hard to deal with" into familiar categories. For instance, a self-driving car that has been taught to stay away from deer, cats, and dogs might not be able to "see" kangaroos crossing the road. A fraud warning algorithm used by the United States Department of Agriculture for the Supplemental Nutrition Assistance Program was trained to detect fraudulent activity on whole-number purchases; however, it failed to detect fraudulent activity at Somali-American grocers, where clients would buy meat in whole dollars. In actuality, the "long tail" can contain items that are not objectively considered "weird": Inappropriate consideration of illnesses like cerebral palsy or diabetes, which are hardly anomalies, was made by the erroneous home health care allocation algorithms.

A multidimensional and cooperative effort among several stakeholders—including developers, regulators, legislators, and civil society organisations— is necessary for a comprehensive strategy to govern AI that accounts for its ethical, social, and economic consequences. Such a strategy should include the following components:

- **Fairness, accountability, transparency, and responsibility** should be given top priority when developing and deploying AI systems. Their design and implementation should respect human rights, embrace diversity, and promote the welfare of society, as recommended by programmers and regulators.

- **Public engagement** is essential to ensure that the needs and issues of all stakeholders are addressed. This may involve consultation with community organisations, civil society organisations, and those who may be impacted by the use of AI systems.
- **Data governance** should be regulated by specific regulations on collection and use, supported by authorisation frameworks that safeguard privacy. Data protection and privacy regulations must be followed by those who build and deploy AI systems.
- **Human oversight** is critical to guarantee that they are operated safely and responsibly. This may entail deploying "human-in-the-loop" technologies, which permit people and AI systems to collaborate on decision-making.
- **Risk assessment and management** should be integral when developing and deploying AI systems. Developers and regulators must assess the ethical, societal, and economic implications of AI systems and implement mitigation plans accordingly.
- **Interdisciplinary collaboration** among experts from a range of fields, including computer science, law, ethics, social sciences, and the humanities, ensures that broader societal impacts of the employment of AI systems are considered.
- **International coordination and cooperation** are necessary to address the global nature of AI. In order to advance ethical and responsible AI, governments and civil society organisations should work together to establish international standards and recommendations that promote ethical and responsible AI.
- Regulatory reform is required, such as amending existing AI regulations (e.g., EU AI Act) to mandate AIAs for all high-risk public-sector AI and private systems in healthcare, hiring, finance, and criminal justice.

By embracing this balanced approach, statutory law can play a crucial role in guiding AI towards a future where technological advancement and societal well-being are not mutually exclusive, but are instead harmoniously integrated. In doing so, we can harness the full potential of AI to drive progress and innovation while upholding our collective responsibility to ethical standards and human values. Ultimately, AI systems should be developed and implemented in ways that benefit all members of society and promote responsible, transparent, and ethical algorithms in line with principles of fairness, public participation, data governance, risk assessment, and international collaboration.

REFERENCES

1. Alam S and others, 'Unmanned and Autonomous Weapons Systems: Practices and Related Policy' (2020) 2(1) PPRI Student Papers in Public Policy 7.
2. Ali W and Hassoun M, 'Artificial Intelligence and Automated Journalism: Contemporary Challenges and New Opportunities' (2019) 5(1) International Journal of Media, Journalism and Mass Communications 40. doi:10.20431/2454-9479.0501004
3. Biresaw SM, 'The Impacts of Artificial Intelligence on Research in the Legal Profession' (2023) 5(1) International Journal of Law and Society 53. doi:10.11648/j.ijls.20220501.17
4. Brynjolfsson E and McAfee A, 'Artificial Intelligence, for Real' (2017) 1 Harvard Business Review 1.
5. Buchanan BG and Headrick TE, 'Some Speculation about Artificial Intelligence and Legal Reasoning' (1970) 23(1) Stanford Law Review 40. doi:10.2307/1227753
6. Buiten MC, 'Towards Intelligent Regulation of Artificial Intelligence' (2019) 10(1) European Journal of Risk Regulation 41.
7. Carrillo MR, 'Artificial Intelligence: From Ethics to Law' (2020) 44(6) Telecommunications Policy 101937. doi:10.1016/j.telpol.2020.101937
8. Carter SM and others, 'The Ethical, Legal and Social Implications of Using Artificial Intelligence Systems in Breast Cancer Care' (2020) 49 The Breast 25. doi:10.1016/j.breast.2019.10.001
9. Castro D and New J, *The Promise of Artificial Intelligence* (Center for Data Innovation 2016)
10. Cath C, 'Governing Artificial Intelligence: Ethical, Legal and Technical Opportunities and Challenges' (2018) 376(2133) Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 20180080. doi:10.1098/rsta.2018.0080
11. Čerka P, Grigienė J and Sirbikytė G, 'Liability for Damages Caused by Artificial Intelligence' (2015) 31(3) Computer Law & Security Review 376. doi:10.1016/j.clsr.2015.03.008
12. Cheatham BM, Javanmardian K and Samandari H, 'Confronting the Risks of Artificial Intelligence' (2019) 2 McKinsey Quarterly 38.
13. Chen B and Chen J, 'China's Legal Practices Concerning Challenges of Artificial General Intelligence' (2024) 13(5) Laws 60. doi:10.3390/laws13050060
14. Chen RJ and others, 'Synthetic Data in Machine Learning for Medicine and Healthcare' (2021) 5(6) Nature Biomedical Engineering 493. doi:10.1038/s41551-021-00751-8
15. Cihon P, 'Standards for AI Governance: International Standards to Enable Global Coordination in AI Research & Development' (*Centre for the Governance of AI (GovAI)*, 17 April 2019)

16. Cobbe J and Singh J, 'Artificial Intelligence as a Service: Legal Responsibilities, Liabilities, and Policy Challenges' (2021) 42 Computer Law & Security Review 105573. doi:10.2139/ssrn.3824736

17. Davenport T and others, 'How Artificial Intelligence Will Change the Future of Marketing' (2020) 48 Journal of the Academy of Marketing Science 24. doi:10.1007/s11747-019-00696-0

18. Davenport TH and Ronanki R, 'Artificial Intelligence for the Real World' (2018) 96(1) Harvard Business Review 108.

19. DeCosta FA, 'Intellectual Property Protection for Artificial Intelligence' [2017] Westlaw Journal Intellectual Property.

20. Duan Y, Edwards JS and Dwivedi YK, 'Artificial Intelligence for Decision Making in the Era of Big Data–Evolution, Challenges and Research Agenda' (2019) 48 International Journal of Information Management 63. doi:10.1016/j.ijinfomgt.2019.01.021

21. Dwivedi YK and others, 'Artificial Intelligence (AI): Multidisciplinary Perspectives on Emerging Challenges, Opportunities, and Agenda for Research, Practice and Policy' (2021) 57 International Journal of Information Management 101994. doi:10.1016/j.ijinfomgt.2019.08.002

22. Edwards LF, *The People and their Peace: Legal Culture and the Transformation of Inequality in the Post-Revolutionary South* (University of North Carolina Press 2014)

23. Erdélyi OJ and Goldsmith J, 'Regulating Artificial Intelligence: Proposal for a Global Solution' (AIES '18: AAAI/ACM Conference on AI, Ethics, and Society, New Orleans LA USA, 2-3, February 2018) 95.

24. Freeman RB, 'Labour Market Institutions Without Blinders: The Debate Over Flexibility and Labour Market Performance' (2005) 19(2) International Economic Journal 129. doi:10.1080/10168730500080675

25. Garcia AR, 'AI, IoT, 'Big Data, and Technologies in Digital Economy with Blockchain at Sustainable Work Satisfaction to Smart Mankind: Access to 6th Dimension of Human Rights' in Lopes NVM (ed), *Smart Governance for Cities: Perspectives and Experiences* (Springer 2020) 83. doi:10.1007/978-3-030-22070-9_6

26. Gerke S, Minssen T and Cohen G, 'Ethical and Legal Challenges of Artificial Intelligence-Driven Healthcare' in Bohr A and Memarzadeh K (eds), *Artificial Intelligence in Healthcare* (Academic Press 2020) 295. doi:10.1016/B978-0-12-818438-7

27. Gershman SJ, Horvitz EJ and Tenenbaum JB, 'Computational Rationality: A Converging Paradigm for Intelligence in Brains, Minds, and Machines' (2015) 349(6245) Science 273. doi:10.1126/science.aac6076

28. Gopalan S and Guihot M, 'Recognition and Enforcement in Cross-Border Insolvency Law: A Proposal for Judicial Gap-Filling' (2015) 48 Vanderbilt Law Review 1225.

29. Greenberg M, 'Legal Interpretation and Natural Law' (2020) 89(1) Fordham Law Review 109.

30. Han X and Althabhawi NM, 'Establishment of Data Intellectual Property Rights and Anti-Monopoly Regulation in China' (2024) 34(2) Jurnal Undang-Undang dan Masyarakat 190. doi:10.17576/juum-2024-3402-13

31. Hanson FA, 'From Key Numbers to Keywords: How Automation Has Transformed the Law' (2002) 94 Law Library Journal 563.

32. Helbing D, 'Societal, Economic, Ethical and Legal Challenges of the Digital Revolution: From Big Data to Deep Learning, Artificial Intelligence, and Manipulative Technologies' in Helbing D (ed), *Towards Digital Enlightenment* (Springer 2019) 47. doi:10.1007/978-3-319-90869-4_6

33. Hoffman S and Podgurski A, 'Artificial Intelligence and Discrimination in Health Care' (2019) 19(3) Yale Journal of Health Policy, Law, and Ethics 1.

34. Jeffries JC, 'Legality, Vagueness, and the Construction of Penal Statutes' (1985) 71(2) Virginia Law Review 189. doi:10.2307/1073017

35. Jordan MI and Mitchell TM, 'Machine Learning: Trends, Perspectives, and Prospects' (2015) 349(6245) Science 255. doi:10.1126/science.aaa841

36. Kaplan A and Haenlein M, 'Rulers of the World, Unite! The Challenges and Opportunities of Artificial Intelligence' (2020) 63(1) Business Horizons 37. doi:10.1016/j.bushor.2019.09.003

37. Koene A and others, *A Governance Framework for Algorithmic Accountability and Transparency* (EU 2019). doi:10.2861/59990

38. Kossow N, Windwehr S and Jenkins M, *Algorithmic Transparency and Accountability* (Transparency International 2021)

39. Kress K, 'Legal indeterminacy' (1989) 77(2) California Law Review 283. doi:10.15779/Z380B17

40. Landes WM and Posner RA, *The Economic Structure of Intellectual Property Law* (Harvard UP 2003)

41. Liu HW, Lin CF and Chen YJ, 'Beyond State v Loomis: Artificial Intelligence, Government Algorithmization and Accountability' (2019) 27(2) International Journal of Law and Information Technology 122.

42. Makridakis S, 'The Forthcoming Artificial Intelligence (AI) Revolution: Its Impact on Society and Firms' (2017) 90 Futures 46. doi:10.1016/j.futures.2017.03.006

43. Mantelero A and Esposito MS, 'An Evidence-Based Methodology for Human Rights Impact Assessment (HRIA) in the Development of AI Data-Intensive Systems' (2021) 41 Computer Law & Security Review 105561. doi:10.1016/j.clsr.2021.105561

44. Marks T, 'John West and the Future of Legal Subscription Databases' (2015) 107(3) Law Library Journal 377. doi:10.2139/ssrn.2441734.

45. Meinzen-Dick RS and Pradhan R, *Legal Pluralism and Dynamic Property Rights* (CAPRi Working Paper no 22, International Food Policy Research Institute 2002)

46. Mitchell M, *Artificial Intelligence: A Guide for Thinking Humans* (Penguin UK 2019)

47. Naik N and others, 'Legal and Ethical Consideration in Artificial Intelligence in Healthcare: Who Takes Responsibility?' (2022) 9 Frontiers in Surgery 266. doi:10.3389/fsurg.2022.862322

48. Nasim SF, Ali MR and Kulsoom U, 'Artificial Intelligence Incidents & Ethics a Narrative Review' (2022) 2(2) International Journal of Technology, Innovation and Management 52. doi:10.54489/ijtim.v2i2.80

49. Paulus JK and Kent DM, 'Predictably Unequal: Understanding and Addressing Concerns that Algorithmic Clinical Prediction May Increase Health Disparities' (2020) 3(1) NPJ Digital Medicine 99. doi:10.1038/s41746-020-0304-9

50. Rai S, 'Legal Liability Issues and Regulation of Artificial Intelligence' (thesis, 2022)

51. Rankin A and others, 'Resilience in Everyday Operations: A Framework for Analyzing Adaptations in high-Risk Work' (2014) 8(1) Journal of Cognitive Engineering and Decision Making 78. doi:10.1177/1555343413498753

52. Re RM and Solow-Niederman A, 'Developing Artificially Intelligent Justice' (2019) 22 Stanford Technology Law Review 242.

53. Risius M and Spohrer K, 'A Blockchain Research Framework: What We (don't) Know, Where We Go from Here, and How We Will Get There' (2017) 59 Business & information systems engineering 385.

54. Robinson PH, 'Fair Notice and Fair Adjudication: Two Kinds of Legality' (2005) 154 University of Pennsylvania Law Review 335.

55. Roco MC and Bainbridge WS, 'The New World of Discovery, Invention, and Innovation: Convergence of Knowledge, Technology, and Society' (2013) 15 Journal of Nanoparticle Research 1946. doi:10.1007/s11051-013-1946-1

56. Rolim MR da S, Dias DM dos S and Filho GNV, 'Regulation of Algorithms in Artificial Intelligence Systems: A Possible Proposal for Brazil?' (2024) 17(2) Contribuciones a Las Ciencias Sociales e4924. doi:10.55905/revconv.17n.2-006

57. Rubin EL, 'Law and Legislation in the Administrative State' (1989) 89(3) Columbia Law Review 369.

58. Samoili S and others, *AI Watch: Defining Artificial Intelligence 2.0* (Publications Office of the EU 2021). doi:10.2760/01990

59. Scherer MU, 'Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies' (2016) 29(2) Harvard Journal of Law & Technology 353. doi:10.2139/ssrn.2609777

60. Solum LB, 'Artificially Intelligent Law' (2019) 1 BioLaw Journal 53. doi:10.15168/2284-4503-351

61. Stachokas G, *The Role of the Electronic Resources Librarian* (Chandos 2019). doi:10.1016/C2018-0-02157-X

62. Strauss PL, 'Legislative Theory and the Rule of Law: Some Comments on Rubin' (1989) 89 Columbia Law Review 427.

63. Sunstein CR, *Legal Reasoning and Political Conflict* (OUP 2018)

64. Sunstein CR, 'On the Expressive Function of Law' (1996) 144(5) University of Pennsylvania Law Review 2021. doi:10.2307/3312647

65. Susskind R and Susskind RE, *Tomorrow's Lawyers: An Introduction to your Future* (OUP 2023)

66. Taeihagh A, 'Governance of Artificial Intelligence' (2021) 40(2) Policy and Society 137. doi:10.1080/14494035.2021.1928377

67. Triantis GG, 'Theory of the Regulation of Debtor-in-Possession Financing, A' (1993) 46 Vanderbilt Law Review 901.

68. Van den Hoven J and others, 'Privacy and Information Technology', *Stanford Encyclopedia of Philosophy* (2014).

69. Veale M and Zuiderveen Borgesius F, 'Demystifying the Draft EU Artificial Intelligence Act' (2021) 22(4) Computer Law Review International 97.

70. Winfield AFT and Jirotka M, 'Ethical Governance is Essential to Building Trust in Robotics and Artificial Intelligence Systems' (2018) 376(2133) Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 20180085. doi:10.1098/rsta.2018.0085

71. Wirtz BW, Weyerer JC and Sturm BJ, 'The Dark Sides of Artificial Intelligence: An Integrated AI Governance Framework for Public Administration' (2020) 43(9) International Journal of Public Administration 818. doi:10.1080/01900692.2020.1749851

72. Yanisky-Ravid S, 'Generating Rembrandt: Artificial Intelligence, Copyright, and Accountability in the 3A Era: The Human-like Authors Are Already Here: A New Model' [2017] Michigan State Law Review 659. doi:10.2139/ssrn.2957722

73. Zhang Y and others, 'Ethics and Privacy of Artificial Intelligence: Understandings from Bibliometrics' (2021) 222(24) Knowledge-Based Systems 106994. doi:10.1016/j.knosys.2021.106994

AUTHORS INFORMATION

Abdesselam Salmi
PhD (Law), Professor of Public law, Ajman University, Ajman, United Arab Emirates
a.salmi@ajman.ac.ae
https://orcid.org/0000-0002-9020-8512
Co-author, responsible for conceptualization, research methodology, writing – original draft, supervising.

Bhupal Bhattacharya
PhD (Law), Assistant Professor, Department of Law, Raiganj University, West Bengal, India
bpb@raiganjuniversity.ac.in
https://orcid.org/0000-0002-7136-3138
Co-author, responsible for research methodology, data collection, writing – original draft.

Sarmistha Bhattacharya
Ph.D. (Law), Assistant Professor & Head, Department of Social Work, AMEX Law College, West Bengal, India
sarmistha.agartala@gmail.com
https://orcid.org/0000-0002-4289-6218
Co-author, responsible for research methodology, data collection, writing – original draft, supervising.

Tarek Abo El Wafa*
Ph.D. (Law), Associate Professor, College of Law, United Arab Emirates University, Al Ain, United Arab Emirates
drtarek@uaeu.ac.ae
https://orcid.org/0000-0002-3923-5187
Corresponding author, responsible for research methodology, data collection, writing – original draft, supervising.

Competing interests: No competing interests were disclosed.

Disclaimer: The authors declare that their opinion and views expressed in this manuscript are free of any impact of any organizations.

RIGHTS AND PERMISSIONS

EDITORS

ABOUT THIS ARTICLE

Cite this article
Salmi A, Bhattacharya B, Bhattacharya S and Abo El Wafa T, 'The Role of Statutory Law in Regulating Artificial Intelligence: Balancing Innovation and Responsibility' (2025) 8(Spec) Access to Justice in Eastern Europe 178-210 <https://doi.org/10.33327/AJEE-18-8.S-a000150>

DOI: https://doi.org/10.33327/AJEE-18-8.S-a000150

Keywords: *AI, limitations of statutory law, technological innovation, policy development, legal frameworks.*

ADDITIONAL INFORMATION

This article is supported by the Ajman University Internal Research Grant No. 2024-IRG-LAW-1. The research findings presented in the paper are the sole responsibility of the authors.

DETAILS FOR PUBLICATION

Date of submission: 24 Apr 2025
Date of acceptance: 17 Jul 2025
Online First publication: 04 Dec 2025
Last Published: 30 Dec 2025
Whether the manuscript was fast tracked? - No
Number of reviewer report submitted in first round: 2 reports
Number of revision rounds: 1 round with major revisions

Technical tools were used in the editorial process:
Plagiarism checks - Turnitin from iThenticate https://www.turnitin.com/products/ithenticate/
Scholastica for Peer Review https://scholasticahq.com/law-reviews

AI DISCLOSURE STATEMENT

AI technologies have only been used to enhance language clarity and grammar. No AI tools were used to generate ideas, structure arguments, analyze data, or produce conclusions.

АНОТАЦІЯ УКРАЇНСЬКОЮ МОВОЮ

Дослідницька стаття

РОЛЬ ЗАКОНОДАВСТВА В РЕГУЛЮВАННІ ШТУЧНОГО ІНТЕЛЕКТУ: БАЛАНС МІЖ ІННОВАЦІЯМИ ТА ВІДПОВІДАЛЬНІСТЮ

Абдесселам Салмі, Бхупал Бхаттахар'я, Сармістха Бхаттахар'я та Тарек Або Ель Вафа*

АНОТАЦІЯ

Вступ. *Штучний інтелект (ШІ) створює серйозні проблеми в управлінні, оскільки його швидка інтеграція в критичні сектори посилює ризики дискримінації, порушення конфіденційності та прогалин у підзвітності. Законодавство, яке традиційно лежить в основі національних правових систем, виявляється дедалі менш придатним для регулювання етичних, соціальних та економічних наслідків ШІ. Його структурна жорсткість у поєднанні з тривалими законодавчими процесами та фрагментацією юрисдикції робить його погано підготовленим до реагування на швидкозмінний характер алгоритмічних технологій. Як наслідок, виникають регуляторні прогалини у сферах застосування з високим рівнем ризику, таких як прогностична поліція, біометричне спостереження, медична діагностика та автономні озброєння, де помилки або упередження можуть призвести до незворотної шкоди. Багато чинних правових норм були створені без урахування складності та непрозорості систем машинного навчання, зокрема їх потенціал функціонувати всупереч традиційним уявленням про людський намір, відповідальність та передбачуваність. Як наслідок, існує нагальна потреба в науковому дослідженні концептуальних та практичних розбіжностей між інноваціями та регулюванням у контексті штучного інтелекту. Також це передбачає вивчення адаптивних правових меж, гібридних моделей управління та інтеграцію етичних принципів у технологічне проєктування.*

Методи. *У статті використовується порівняльно-правовий аналіз нормативно-правового регулювання у сфері штучного інтелекту в ключових юрисдикціях (ЄС, США, Китай, Бразилія, Велика Британія) у поєднанні з доктринальним дослідженням законодавчих текстів та судової практики. Методологія інтегрує систематичний огляд первинних джерел (наприклад, Закон ЄС про ШІ, проєкти Закону США про алгоритмічну підзвітність, тимчасові заходи GenAI Китаю), якісну оцінку вторинної літератури та*

інституційних звітів, використання принципу «Питання-Правило-Застосування-Висновок» для оцінки ефективності регулювання, а також міжюрисдикційне дослідження механізмів правозастосування та стандартів відповідальності.

Результати та висновки. *Аналіз виявляє критичні обмеження статутного права, юрисдикційні розбіжності в класифікації ризиків (наприклад, попередня оцінка відповідності в ЄС проти секторального постфактумного правозастосування в США), фрагментацію відповідальності та прогалини у правозастосуванні. Найголовніше, що самі по собі статутні підходи не можуть збалансувати просування інновацій з етичними обмеженнями: надмірне регулювання гальмує дослідження та розробки, а слабке – сприяє заподіянню шкоди суспільству. У дослідженні зроблено висновок, що ефективне управління вимагає додаткових етичних меж, які впроваджують прозорість, аудит упередженості та людський нагляд; міжнародну гармонізацію стандартів відповідальності та протоколів ризиків; адаптивні регуляторні умови для реальних випробувань; та співпрацю з багатьма зацікавленими сторонами для розробки контекстно-залежних впроваджень.*

Ключові слова: *ШІ, обмеження статутного права, технологічні інновації, розробка політики, правові межі.*

ABSTRACT IN ARABIC

مقال بحثي

دور التشريعات القانونية في تنظيم الذكاء الاصطناعي: الموازنة بين الابتكار والمسؤولية

إعداد: عبد السلام سلمي، وبوبال بهاتاشاريا، وسارميستا بهاتاشاريا، وطارق أبو الوفا*

الملخص

الخلفية: يُشكّل الذكاء الاصطناعي تحديًا عميقًا لأنظمة الحوكمة، إذ إن تسارُع دمجه في القطاعات الحيوية يزيد من احتمالات التمييز، وانتهاك الخصوصية، واتساع فجوات المساءلة القانونية. وقد تبيّن أن القوانين التشريعية التقليدية، التي تشكّل الركيزة الأساسية للأنظمة القانونية الوطنية، أصبحت عاجزة على نحو متزايد عن مواكبة الآثار الأخلاقية والاجتماعية والاقتصادية المترتبة على تطور الذكاء الاصطناعي. فطبيعة هذه القوانين المتسمَة بالجمود البنيوي وبطء العملية التشريعية، إلى جانب تشتّت الأطر القانونية بين السلطات القضائية المختلفة، تجعلها غير قادرة على الاستجابة السريعة للتغير المتسارع في التقنيات الخوارزمية. وينتج عن ذلك فراغ تنظيمي واضح في مجالات عالية

الخطورة مثل التنبؤ الجرمي، والمراقبة البيومترية، والتشخيص الطبي، والأسلحة ذاتية التشغيل، وهي مجالات قد يؤدي فيها الخطأ أو الانحياز إلى أضرار لا يمكن تداركها. كما أن كثيرًا من القواعد القانونية الحالية وُضعت في زمن لم يتخيّل فيه المشرّع التعقيد والغموض الذي تتسم به أنظمة التعلّم الآلي، ولا قدرتها على العمل بطرق تتجاوز المفاهيم القانونية التقليدية حول النية البشرية، والمسؤولية، والتوقّع المسبق. لذلك، تبرز حاجة ملحّة إلى مزيد من البحث الأكاديمي المتعمّق لفهم التوتر القائم بين الابتكار والتنظيم القانوني في سياق الذكاء الاصطناعي، من خلال استكشاف أطر تشريعية أكثر مرونة، ونماذج حوكمة هجينة تجمع بين القانون والتقنية، مع دمج المبادئ الأخلاقية في تصميم الأنظمة الذكية منذ مراحلها الأولى.

المنهجية: اعتمدت هذه الدراسة على تحليل قانوني مقارن للأطر التشريعية المنظِّمة للذكاء الاصطناعي في عدد من الولايات القضائية الرئيسية تشمل الاتحاد الأوروبي، والولايات المتحدة، والصين، والبرازيل، والمملكة المتحدة، إلى جانب بحث فقهي تحليلي للنصوص التشريعية والأحكام القضائية ذات الصلة. تقوم المنهجية على مراجعة منهجية شاملة للمصادر الأولية مثل: قانون الذكاء الاصطناعي الأوروبي (EU AI Act)، ومسودات قانون المساءلة الخوارزمية الأمريكي (Algorithmic Accountability Act)، والتدابير المؤقتة الصينية لتنظيم الذكاء الاصطناعي التوليدي (GenAI Interim Measures)، إضافة إلى تقييم نوعي للمراجع الثانوية والتقارير المؤسسية التي تناولت قضايا الحوكمة والتنظيم. كما استخدم الباحثون إطار تحليل "القضية – القاعدة – التطبيق – النتيجة" (Issue–Rule–Application–Conclusion) لتقدير فعالية الأطر التنظيمية، مع مقارنة عابرة للأنظمة القانونية بغرض تحديد أوجه الاختلاف في آليات التنفيذ ومعايير المسؤولية القانونية بين الدول محل الدراسة.

النتائج والاستنتاجات: أظهرت نتائج التحليل أن القانون التشريعي يواجه قيودًا جوهرية في قدرته على مواكبة تحديات الذكاء الاصطناعي، وأن هناك تباينًا ملحوظًا بين الأنظمة القانونية المختلفة في تصنيف المخاطر وآليات التنظيم. فعلى سبيل المثال، يعتمد الاتحاد الأوروبي نهجًا استباقيًا (وقائيًا) يقوم على تقييم المطابقة المسبق للمخاطر، في حين تتبع الولايات المتحدة أسلوبًا لاحقًا (علاجيًا) يقوم على تطبيق القوانين القطاعية بعد وقوع الانتهاك. من المهم التأكيد على أن الاعتماد على الأطر التشريعية وحدها لا يحقق التوازن المطلوب بين تشجيع الابتكار وضبط الجوانب الأخلاقية؛ فالتنظيم المفرط يؤدي إلى خنق البحث والتطوير وإبطاء وتيرة التقدّم التقني، في حين أن التساهل في القواعد القانونية يفتح الباب أمام أضرار اجتماعية جسيمة تمس العدالة والمساءلة والثقة العامة في التكنولوجيا. خلصت الدراسة إلى أن الحوكمة الفعّالة للذكاء الاصطناعي تتطلب اعتماد أطر أخلاقية مكمّلة تُدمج فيها مبادئ الشفافية، ومراجعة الانحيازات، والإشراف البشري المستمر، إلى جانب العمل على توحيد المعايير الدولية الخاصة بالمسؤولية القانونية وبروتوكولات تقييم المخاطر. كما توصي الدراسة بإنشاء بيئات تنظيمية تجريبية مرنة (Regulatory Sandboxes) تسمح باختبار الأنظمة في ظروف واقعية قبل تطبيقها على نطاق واسع، مع تعزيز التعاون بين مختلف الأطراف المعنية — من حكومات ومطورين وباحثين ومؤسسات مجتمع مدني — لضمان تصميم حلول تنظيمية تراعي خصوصية السياقات القانونية والاجتماعية في كل دولة.

Access to Justice in Eastern Europe
ISSN 2663-0575 (Print)
ISSN 2663-0583 (Online)
http://ajee-journal.com

Research Article

ARTIFICIAL INTELLIGENCE IN CRIMINAL JUSTICE IN GERMANY AND UKRAINE: A COMPARATIVE LEGAL STUDY

Lidiia Moskvych*, Iryna Borodina and Olga Ovsiannikova

DOI:
https://doi.org/10.33327/AJEE-18-8.S-a000155

Date of submission: 03 Sep 2025
Date of acceptance: 01 Oct 2025
Date of Publication 30 Dec 2025

Disclaimer:

The authors declare that opinion and views expressed in this manuscript are free of any impact of any organizations.One of the authors additionally declares that her opinion and views expressed in this manuscript are free from any influence of any organization, including the Constitutional Court of Ukraine, despite the fact that she is a member of the Scientific Advisory Council of the Constitutional Court.

ABSTRACT

Background: *Artificial intelligence (AI) is rapidly evolving from peripheral administrative tools into applications that directly influence the functioning of criminal justice systems. In Europe, this integration proceeds under a cautious, law-centred approach that seeks to balance innovation with the preservation of judicial independence, fairness, and the rule of law. This article offers a comparative legal analysis of AI deployment in the criminal justice systems of Germany and Ukraine, situating national developments within the broader framework of the EU Artificial Intelligence Act (2024), Council of Europe standards, and constitutional safeguards. Germany's structured, federally coordinated rollout contrasts with Ukraine's targeted yet ethically constrained implementation, reflecting divergent institutional capacities and legal traditions.*

Methods: *This study adopts a comparative legal approach that combines functionalist and contextualist perspectives. The functionalist dimension examines how Germany and Ukraine employ AI in criminal justice to address analogous issues—such as efficiency, transparency, and rights protection—while the contextualist dimension situates these developments within each country's constitutional framework, institutional capacity, and socio-political environment, notably the impact of wartime conditions in Ukraine. This combined perspective ensures that similarities and divergences are assessed not in abstraction but against the broader background of European and national legal cultures. The analysis draws on primary law, regulatory instruments, official court and ministerial reports, and peer-reviewed scholarship. Empirical examples include German pilot projects in predictive policing (PRECOBS, KLB-operativ), investigative filtering tools, and administrative AI in courts, as well as Ukraine's probation risk-assessment algorithm Cassandra and AI-assisted systems for legal research and translation. Experiences from the United States with algorithmic risk assessment are used as a cautionary benchmark.*

Results and Conclusions: *The study finds that, while both jurisdictions restrict the use of AI as a substitute for core judicial decision-making, Germany leverages its infrastructure, coordinated administration, and legislative oversight to test and evaluate AI tools. By contrast, Ukraine's integration is more selective, subject to explicit ethical limitations, but hindered by gaps in transparency and the constraints imposed by wartime conditions. The analysis identifies common challenges—including algorithmic bias, explainability, evidentiary admissibility, and the protection of fair trial guarantees—and formulates context-specific recommendations. These include mandatory external audits, codified procedural rights to challenge AI-generated data, clearer evidentiary protocols, and enhanced judicial awareness of AI technologies. The study underscores that the sustainable integration of AI into criminal justice must remain supportive, auditable, and under human control to comply with European legal standards and safeguard fundamental rights.*

1 INTRODUCTION

Across the world, emerging technologies are transforming judicial systems. Countries such as China, Singapore, and the United States have launched pilot projects for "intelligent courts" or algorithmic tools designed to optimise the handling of specific cases, reportedly leading to greater efficiency. In China, for instance, AI-based systems are already assisting in analysing evidence and even drafting decisions in minor civil disputes, substantially reducing case-processing times.[1]

By contrast, European jurisdictions have pursued a more cautious approach, limiting artificial intelligence to supportive functions such as electronic filing systems and digital

1 Changqing Shi, Tania Sourdin and Bin Li, 'The Smart Court – A New Pathway to Justice in China?' (2021) 12(1) International Journal for Court Administration 4. doi:10.36745/ijca.367.

databases, and firmly preserving human judgment in adjudication.[2] A notable expression of this cautious philosophy is the adoption of the European Union's Artificial Intelligence Act (2024),[3] which embodies the regional consensus that technological innovation must be reconciled with reliable oversight and the protection of fundamental rights.

Germany and Ukraine exemplify this model of cautious integration within Europe, albeit under very different circumstances. Germany, a long-standing EU member with a federal judicial system and stringent data protection rules, has implemented a systematic "*Justice 4.0*" strategy. Ukraine, by contrast, as an EU candidate state with a judiciary in transition, is modernising rapidly through the digitalisation of court services despite the obstacles posed by the ongoing armed conflict. For more than a decade, Ukrainian courts have used digital platforms to provide public access to judicial decisions and facilitate remote hearings—measures that have improved transparency and reduced backlogs. Nevertheless, Ukraine's ability to introduce advanced AI technologies remains constrained by limited resources and wartime priorities.

Taken together, these two countries illustrate a striking contrast: one relies on substantial resources and a methodical, EU-based approach, while the other introduces innovations under pressure, guided largely by European recommendations.

2 METHODOLOGY

This study employs a comparative legal methodology designed to illuminate both convergences and divergences in the integration of artificial intelligence within the criminal justice systems of Germany and Ukraine. The analysis rests on three interrelated pillars.

First, a functionalist inquiry identifies how AI technologies are deployed to address analogous challenges—efficiency in case management, transparency in judicial reasoning, and the safeguarding of fundamental rights. This dimension enables the systematic comparison of institutional responses to shared problems without presuming that similar technologies necessarily yield identical legal consequences.

Second, a contextualist approach situates these functional developments within each jurisdiction's constitutional architecture, institutional capacity, and socio-political environment. Particular attention is given to the embeddedness of AI regulation within

2 Elif Kiesow Cortez and Nestor Maslej, 'Adjudication of Artificial Intelligence and Automated Decision-Making Cases in Europe and the USA' (2023) 14(3) European Journal of Risk Regulation 457. doi:10.1017/err.2023.61.

3 Regulation (EU) 2024/1689 of the European Parliament and of the Council 'Laying Down Harmonised Rules on Artificial Intelligence and Amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)' (13 June 2024) <https://eur-lex.europa.eu/eli/reg/2024/1689/oj> accessed 10 August 2025.

European legal culture in Germany, and to the constraints and innovations shaped by wartime conditions in Ukraine. By embedding national practices within their broader normative and political contexts, the analysis avoids decontextualised parallels and instead highlights the distinctive logics guiding adoption.

Third, a multi-source research design provides a robust evidentiary base for the comparative exercise. The study draws on primary legislation, constitutional and international instruments, ministerial and judicial policy documents, and case law, complemented by secondary materials such as peer-reviewed scholarship and professional reports. Empirical illustrations are incorporated through pilot projects and operational tools, including German initiatives in predictive policing (PRECOBS, KLB-operativ), investigative data-filtering algorithms, and administrative AI systems in courts; and Ukraine's Cassandra risk-assessment program and AI-supported legal research and translation services. Comparative references to the United States' experience with algorithmic risk assessment further serve as a cautionary benchmark.

Analytically, the study combines doctrinal legal reasoning with normative evaluation. Doctrinally, it maps the extent to which AI applications comply with constitutional safeguards, data protection requirements, and procedural rights. Normatively, it interrogates whether these safeguards are adequate to sustain fairness, accountability, and human oversight in criminal justice. This dual perspective allows the inquiry to transcend mere description and to formulate context-sensitive recommendations.

In sum, the methodology integrates functionalist comparison, contextualist interpretation, and multi-source triangulation to provide a comprehensive and rigorous assessment of how two distinct legal systems navigate the opportunities and risks of AI in criminal justice.

3 LEGAL AND POLITICAL FOUNDATIONS IN EUROPE

Before examining national developments, it is necessary to outline the general European legal and ethical framework governing the use of AI in the administration of justice. Both Germany and Ukraine operate within this framework, which establishes fundamental restrictions and requirements for any application of AI in criminal proceedings.

At its core lies the EU Artificial Intelligence Act (2024),[4] which establishes a comprehensive risk-based regulatory framework. Crucially, it classifies AI systems deployed by law enforcement or judicial authorities as "high-risk," thereby imposing stringent obligations. Providers and users of such systems must implement documented risk-management processes, ensure the high quality of training data, guarantee human oversight at critical decision points, maintain detailed operational logs, and provide transparency to affected individuals. Certain AI practices are categorically prohibited as posing an "unacceptable

4 ibid.

risk"—for example, social scoring and real-time biometric identification in public spaces—because they are incompatible with fundamental rights. The designation of judicial and law-enforcement AI as "high-risk" means that tools such as predictive policing software or decision-support systems must undergo prior conformity assessments and continuous monitoring to ensure compliance. The EU Artificial Intelligence Act entered into force on 1 August 2024 and will become fully applicable on 2 August 2026. Germany, as an EU member, will apply its provisions directly, while Ukraine, as a candidate state, is preparing its legislation to align with them.

The Council of Europe has likewise issued influential ethical recommendations on the use of AI in judicial systems, which are recognised by both Germany and Ukraine. The European Ethical Charter on the Use of AI in Judicial Systems[5] (adopted by CEPEJ in 2018) identifies five guiding principles: (1) respect for fundamental rights; (2) non-discrimination; (3) quality and security; (4) transparency and impartiality; and (5) user control. In essence, these principles require that any application of AI respect the rights guaranteed by the European Convention on Human Rights, avoid perpetuating bias or injustice, function reliably and securely, remain open to scrutiny—both in terms of its operation and in ensuring that judicial impartiality is not undermined—and remain subject to the final authority of human judicial actors.

In 2024, the Council of Europe also opened for signature the world's first binding international treaty on AI: the Framework Convention on AI and Human Rights.[6] This instrument obliges signatory states to adopt safeguards proportionate to the risks posed by AI systems, ensure effective remedies for individuals affected by AI-based decisions, and preserve judicial independence and the rule of law in the era of automation. Although neither Germany nor Ukraine has yet ratified this newly established convention, its principles reinforce those of the Ethical Charter. Taken together, the EU and Council of Europe instruments establish a common foundation: AI may be used to enhance efficiency and access to justice, but it must not compromise fundamental rights, equality, or the human-centred character of the judiciary.

Beyond the European legal framework, global ethical and policy debates provide additional critical perspectives. UNESCO's Recommendation on the Ethics of Artificial Intelligence (2021)[7] and its ongoing Judicial AI Ethics Guidelines Project (2025)[8] underline concerns

5 CEPEJ, *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their Environment* (Council of Europe 2018) <https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c> accessed 10 August 2025.

6 Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (5 September 2024) CETS 225 <https://rm.coe.int/1680afae3c> accessed 10 August 2025.

7 UNESCO, *Recommendation on the Ethics of Artificial Intelligence: Adopted on 23 November 2021* (SHS/BIO/PI/2021/1, UNESCO 2022) <https://unesdoc.unesco.org/ark:/48223/pf0000381137> accessed 10 August 2025.

8 UNESCO, *Draft Guidelines for the Use of AI Systems in Courts and Tribunals* (CI/DIT/2025/GL/01, UNESCO 2025) <https://unesdoc.unesco.org/ark:/48223/pf0000393682> accessed 10 August 2025.

about fairness, transparency, and human oversight. Interdisciplinary scholarship on AI ethics and digital governance[9] similarly emphasises accountability gaps, opacity, and the societal risks of overreliance on algorithms. Integrating these perspectives situates the experiences of Germany and Ukraine within a broader global conversation on responsible AI, highlighting that the challenges at stake are not only legal but also ethical and societal. Complementing these debates, the OECD/G20 AI Principles (2019)[10] further underline global consensus on human-centred values, transparency, robustness, and accountability.

These principles—endorsed by leading economies—situate the European and national experiences of Germany and Ukraine within a truly international governance landscape.

4 GERMANY: USE CASES, REGULATION, AND DEVELOPMENT PATHWAYS

Germany has adopted a proactive yet cautious stance towards the integration of AI into its judicial system, characterised by close coordination between federal and state authorities (Bund und Länder). In June 2025, for example, the justice ministers of the federal government and all 16 Länder issued a Joint Declaration on the Use of AI in the Judiciary, setting the tone for future adoption.[11] The declaration commits the German judiciary to employing AI in a manner that is "responsible, comprehensible, and reliable," highlighting its potential to enhance efficiency in routine tasks and in handling cases with substantial information volumes, while making it unequivocally clear that AI is not to replace judges or judicial discretion. In other words, every algorithm is to remain a tool for human decision-makers rather than a decision-maker in itself. The joint strategy also emphasises the importance of transparency and accountability for any AI deployed within the courts.

This high-level policy is reinforced by concrete national initiatives. The federal Digital Summit Communiqué introduced the modernisation program *Justice 4.0*, which envisions the development of a nationwide platform for court translations (to support multilingual proceedings), the exploration of a large language model (LLM) owned by the judiciary for legal research, and the creation of a unified IT architecture for the justice system. A tangible outcome is the establishment of a secure nationwide judicial cloud,

9 Luciano Floridi and Josh Cowls, 'A Unified Framework of Five Principles for AI in Society' (2019) 1(1) Harvard Data Science Review 1. doi:10.1162/99608f92.8cd550d1; Brent Mittelstadt, 'Principles Alone Cannot Guarantee Ethical AI' (2019) 1 Nature Machine Intelligence 501. doi:10.1038/s42256-019-0114-4.

10 G20, 'G20 Ministerial Statement on Trade and Digital Economy; Annex: G20 AI Principles' (9 June 2019) <https://oecd.ai/en/wonk/documents/g20-ai-principles> accessed 10 August 2025; OECD, *Recommendation of the Council on Artificial Intelligence* (OECD/LEGAL/0449, OECD Legal Instruments 2025) <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449> accessed 10 August 2025.

11 Justizminister von Bund und Ländern, 'Gemeinsame Erklärung zum Einsatz von KI in der Justiz' (June 2025) <https://justiz.de/laender-bund-europa/bund_laender/Gemeinsame-Erklaerung-zum-Einsatz-KI/index.php;jsessionid=21C24681D0563738863495FD5BA330C2> accessed 10 August 2025.

approved at the end of 2024, designed to host judicial IT services (including future AI tools) within a standardised and secure environment. The judicial cloud is intended to ensure that all AI systems used in German courts remain under national jurisdiction and comply with rigorous data-security protocols.[12]

Within the field of criminal justice, several AI-based tools have been piloted or deployed, though predominantly under the auspices of police forces and judicial administrations rather than by judges in courtrooms. These applications remain limited in scope and are subject to careful evaluation.

First, Germany was among the earliest European jurisdictions to experiment with predictive policing software. Several Länder have piloted systems that analyse crime data to forecast where offences are likely to occur. Bavaria, for example, has employed the PRECOBS (Pre-Crime Observation System) tool to anticipate burglaries; the police in Hesse have used a system known as KLB-operativ; and Berlin briefly tested an instrument called KrimPro.[13] These programs process historical crime reports and other datasets to detect patterns (such as locations repeatedly targeted by theft) and generate "risk maps" or alerts for law enforcement agencies. In practice, their use has been uneven and their outcomes inconclusive. Baden-Württemberg, for instance, discontinued the testing of PRECOBS, judging its predictive value insufficient to justify the costs and data demands, whereas Bavaria continues to employ it in a limited capacity to optimise patrol distribution for burglary prevention. Preliminary assessments indicate that, although these tools may correlate with modest reductions in property crime in targeted areas, their impact on overall crime rates has been marginal, while concerns persist regarding data quality and "blind spots."[14] It is noteworthy that such predictive systems are used at the investigative stage, prior to court proceedings, but may nonetheless indirectly affect the course of criminal justice by shaping police focus, arrests, and subsequent prosecutions.[15]

Second, German law enforcement authorities are increasingly deploying AI-based tools to filter and analyse vast datasets in criminal investigations. The growing prevalence of digital evidence—from mobile phone data and surveillance camera footage to online communications—has rendered manual review impracticable. AI algorithms are thus used to flag relevant images or messages (for example, in child abuse cases) and to monitor open-

12 Nicola Hauptmann, 'Aufbau der bundesweiten Justiz-Cloud beschlossen' (*eGovernment: Verwaltung Digital*, 9 December 2024) <https://www.egovernment.de/aufbau-der-bundesweiten-justizcloud-beschlossen-a-782965/> accessed 10 August 2025.

13 'PRECOBS - Predictive Policing in German Administrations' (*IPS-X*, 2018) <https://ipsoeu.github.io/ips-explorer/case/10433.html> accessed 10 August 2025.

14 Libuše Hannah Vepřek and others, 'Legitimising Predictive Policing in Germany' (2020) 2(3) Kriminologie 1. doi:10.18716/ojs/krimoj/2020.3.3.

15 Amelie Spell, 'The Use of Predictive Policing in German Law Enforcement: A Discourse Analysis' (Bachelor's thesis, University of Twente 2023) <https://purl.utwente.nl/essays/96893> accessed 10 August 2025.

source information (OSINT) from social networks during investigations. While these tools hold promise for the efficient processing of "big data," German scholars and practitioners have highlighted significant evidentiary implications. Key concerns include explainability (investigators must understand why the AI flagged particular elements), the integrity of the chain of custody (ensuring that AI processing does not distort or invalidate evidence), and the risk of confirmation bias (where investigators may give disproportionate weight to AI-identified items). In light of these concerns, internal guidelines often require human analysts to review AI outputs, and any new forensic algorithm must undergo legal vetting for compliance with evidentiary rules.[16]

Third, within the judiciary, AI adoption has so far been concentrated in civil proceedings, notably in the automated processing of mass debt-collection cases. In criminal justice, the use of AI by judges or prosecutors remains largely experimental. Nonetheless, the Federal Ministry of Justice's Digitalisation Fund is investing in AI initiatives (*KI-Vorhaben*) that could eventually be extended to criminal proceedings. Current projects include tools for automating the anonymisation of court decisions—a prerequisite for publication—and real-time translation of court hearings. Such instruments could ultimately prove valuable in criminal courts, particularly in cases with international dimensions or sensitive personal data. Additionally, prosecutors are piloting text-analysis algorithms designed to sort and summarise voluminous case materials, such as those arising in complex financial crime investigations.

It is important to stress that no German court or correctional authority employs algorithmic risk-assessment tools in sentencing, bail determinations, or parole decisions—a deliberate departure from U.S. practice.[17] The judiciary has categorically rejected the use of "risk scores" in adjudication, reflecting a legal culture that demands individualised, reasoned judgments by judges and remains sceptical of opaque indicators that could lead to deprivations of liberty.[18] As one observer has noted, criminal justice in Germany remains "judge-centred": while administrative tasks may be optimised through technology, the act of judicial decision-making itself remains insulated from automation.[19]

16 Johanna Sprenger and Dominik Brodowski, '"Predictive Policing", "Predictive justice", and the Use of AI in the Administration of Justice in Germany' [2023] e-Revue Internationale de Droit Pénal 117. doi:10.22028/D291-39980.

17 Case No 2015AP157-CR *State of Wisconsin v Eric L Loomis* (Wisconsin Supreme Court, 13 July 2016) <https://www.wicourts.gov/sc/opinion/DisplayDocument.pdf?content=pdf&seqNo=171690> accessed 10 August 2025.

18 Ministerium der Justiz des Landes Nordrhein-Westfalen and Landtag Nordrhein-Westfalen, Kooperationsvereinbarung und Werkvertrag für das Vorhaben Generatives Sprachmodell der Justiz (GSJ) im Rahmen der Digitalisierungsinitiative für die Justiz (Drucksache 18/2717, 20 June 2024) <https://www.landtag.nrw.de/portal/WWW/dokumentenarchiv/Dokument/MMV18-2717.pdf> accessed 10 August 2025.

19 'Justiz: KI soll Richter nicht ersetzen' (*Move-online*, June 2023). <https://www.move-online.de/> accessed 10 August 2025.

Any use of AI within the German justice system is also subject to stringent legal constraints. The German Basic Law (Grundgesetz) enshrines human dignity and the rule of law, which, according to judicial interpretation, prohibit fully automated decision-making in areas that significantly affect individual rights. Combined with Germany's obligations under the European Convention on Human Rights—such as the right to a fair trial and equality before the law—this constitutional backdrop ensures that human judges must remain responsible for sentencing and adjudication. In addition, German data protection law imposes further restrictions: Article 22 of the EU General Data Protection Regulation (GDPR),[20] implemented domestically through the Federal Data Protection Act (BDSG),[21] prohibits decisions that produce legal effects for individuals from being based solely on automated processing, except in narrowly defined circumstances. This means that if an AI system is used to assist decision-making in a criminal case, a human official must carefully review and approve the outcome rather than merely rubber-stamp the algorithm's result. The use of profiling or predictive analytics is further subject to the GDPR's principles of necessity and proportionality.

Beyond these binding legal requirements, Germany has also issued political recommendations on the deployment of AI. The 2025 Joint Declaration of the Federal Government and the Länder sets out clear red lines: AI tools may be used to enhance efficiency in routine processes, but "in all cases the independence of the judge and his or her decision-making authority must remain inviolable," and all AI outputs must be explainable and verifiable.[22] The declaration and related policy documents also foresee accountability mechanisms, including documentation and audit trails for each AI system used. Meanwhile, ethics councils and professional associations (including judicial associations) are drafting guidelines for AI. These include recommendations that every algorithmic tool should be tested for bias prior to use, that judges should be trained in the limitations of AI, and that defendants or counsel should be informed if AI has been used to analyse evidence in their cases. Although these guidelines are not yet legally binding, they reflect Germany's strong inclination toward caution in the use of AI in criminal justice—an effort to gradually increase efficiency while rigorously safeguarding constitutional rights.

20 Regulation (EU) 2016/679 of the European Parliament and of the Council 'On the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation)' (27 April 2016) [2016] OJ L 119/1 <http://data.europa.eu/eli/reg/2016/679/oj> accessed 10 August 2025.

21 Federal Data Protection Act (BDSG) (30 June 2017) <https://www.gesetze-im-internet.de/englisch_bdsg/> accessed 10 August 2025.

22 Justizminister von Bund und Ländern, 'Gemeinsame Erklärung (n 11).

5 UKRAINE: USE CASES, REGULATION, AND DEVELOPMENT PATHWAYS

Ukraine's judicial and legal system has demonstrated openness to technological innovation while deliberately establishing ethical boundaries from the outset. In September 2024, the Judicial Council of Ukraine amended the national Code of Judicial Ethics by introducing a new Article 16 addressing the use of AI.[23] This provision permits judges to employ AI tools in their work only under strict conditions: the use of AI must neither compromise judicial independence or impartiality, nor affect the evaluation of evidence or the substance of judicial decisions, nor breach existing laws.

In effect, Ukraine has codified the principle of the "exclusively supportive use" of AI in adjudication. Judges may, for instance, utilise an AI-powered translation service or a tool for legal research, but they are expressly prohibited from delegating to AI the assessment of evidence or the determination of guilt or innocence. This position has been underscored in public statements by the judiciary, which has repeatedly affirmed that technology will not replace human reasoning. Commentators note that this approach is consistent with European ethical standards and helps reinforce public trust during a period of rapid transformation.[24]

Ukrainian legal scholars and reformers are actively debating the future integration of AI into the judicial system. Many argue that rules and legislation must be formalised now—before the widespread adoption of AI—in order to prevent ethical violations. For example, Zhukevych, Moskvych, Manhora et al. emphasise the growing importance of establishing a comprehensive legal framework for the use of AI in judicial proceedings and law enforcement, underscoring that the introduction of such technologies must be accompanied by clear legal safeguards.[25] Others highlight the potential benefits of AI. A striking example is the empirical study by Izarova and colleagues, which explored the application of AI-driven precedent analysis to the Unified State Register of Court Decisions.[26] This national archive, containing more than 100 million judicial decisions, has

23 Decision of the Congress of Judges of Ukraine 'On Approval of the Code of Judicial Ethics' of 22 February 2013 (amended 18 September 2024) <https://zakon.rada.gov.ua/rada/show/n0001415-13#Text> accessed 10 August 2025.

24 Ian Bernaziuk, 'Artificial Intelligence and the Judicial System of Ukraine: Results of Cooperation in the Past Year: Presentation' (JAR-Association Conference "Judicial Systems in Transition: Reforms, Innovations and Justice", JAR-Association, OMIJ, Faculty of Law University of Limoges, 11 June 2025) <https://court.gov.ua/storage/portal/supreme/prezentacii_2025/AI_Ukraine_bernaziuk.pdf> accessed 10 August 2025; Yuliia Moskvytyn and Agne Limante, 'Integrating Artificial Intelligence in Ukraine's Courts: State of Play and Future Prospect' (*Verfassungsblog*, 12 December 2024) <https://verfassungsblog.de/ai-ukraine-judiciary/> accessed 10 August 2025.

25 Ihor Zhukevych and others, 'Analysis of Issues Related to the Legalization of Artificial Intelligence, Its Use in Legal Proceedings, Legal Consultation and Law Enforcement' (2024) 27 Legal Scientific Journal 17.

26 Iryna Izarova and others, 'Advancing Sustainable Justice Through AI-Based Case Law Analysis: Ukrainian Experience' (2024) 7(1) *Access to Justice in Eastern Europe* 127. doi:10.33327/AJEE-18-7.1-a000123.

proven to be an unparalleled resource for identifying systemic deficiencies, developing predictive models for case outcomes, and providing an objective empirical foundation for judicial reform. These discussions frequently point out that Ukraine's aspiration to join the European Union provides a strong incentive to align national legislation with EU trends in AI governance, while simultaneously adapting legal solutions to the post-Soviet legal culture and the current realities of the country.

Despite difficult circumstances, Ukraine has already introduced several noteworthy applications of AI in criminal justice. One of the earliest was the 2020 launch of Cassandra, an algorithm developed by the State Probation Service under the Ministry of Justice[27] to predict the risk of recidivism among convicted offenders eligible for probation or early release. Probation officers input data from structured questionnaires covering the offender's biography, the nature of the offence, and behavioural characteristics, after which the system generates a risk assessment intended to guide decisions on the level of supervision or recommendations to courts regarding release. Officials describe the system as a hybrid decision-support tool that combines human judgment with algorithmic suggestions.[28]

However, limited public information is available regarding *Cassandra's* internal operation or accuracy. Neither the full list of input variables nor evaluations of its prediction accuracy have been published. This lack of transparency has prompted criticism from civil society and academia, who express concern about possible biases or errors in the algorithm that could influence judicial decision-making. Questions remain unanswered: Does Cassandra disproportionately classify certain groups as high-risk? How does it account for socio-economic factors? The absence of transparency reinforces calls for greater openness concerning the tool's design and operation.[29]

The judiciary has likewise embraced AI to enhance its capacity for legal research and to overcome linguistic barriers. In 2025, the Supreme Court of Ukraine modernised its Unified Legal Positions database by incorporating an AI-based search engine and even experimental generative AI functions designed to assist judges and clerks in rapidly locating relevant precedents and summaries of legal principles. The system employs natural language processing, allowing users to submit queries in plain Ukrainian (or potentially in English) and receive concise answers or lists of relevant cases—resembling, in some respects, an advanced legal chatbot. In addition, the Supreme Court has developed an AI-powered translation service for judicial documents and decisions, including a module trained on previous translations of European Court of Human Rights judgments to ensure accuracy in

27 Fair Trials, *Automating Injustice: The Use of Artificial Intelligence & Automated Decision-Making Systems in Criminal Justice in Europe* (Fair Trials 2021) <https://www.fairtrials.org/articles/publications/automating-injustice/> accessed 10 August 2025.

28 Johanna Jacobson, 'Algorithmic Risk Assessment Tools in Criminal Proceedings: An Analysis in Light of Articles 6 and 14 of the European Convention on Human Rights' (Master's thesis, Uppsala University 2022).

29 Moskvytyn and Limante (n 24).

Ukrainian legal terminology. These innovations are clearly designed to improve productivity: they save time in legal research and ensure that parties receive information in their own language, but they play no role in judicial decision-making itself. Their rollout was accompanied by training sessions for judges, and early reports reflect cautious enthusiasm: judges value the assistance in navigating vast jurisprudential archives, yet remain acutely aware that the responsibility for legal reasoning rests exclusively with them.[30]

In contrast to Germany, Ukraine has not yet introduced predictive policing or other AI-driven investigative methods on a significant scale. This is partly due to limited resources and institutional capacity, partly due to the ongoing war, which has shifted national priorities. Nonetheless, Ukrainian law enforcement has been exploring the potential of AI in specific areas. Pilot projects have investigated the use of machine learning to match ballistic evidence or to detect suspicious financial transactions linked to corruption. Academic discussions have also considered the possible use of facial recognition or video analytics for suspect identification and public security monitoring. Yet, as of 2025, such ideas remain at the stage of pilot projects or proposals—publicly available information indicates that the Ukrainian police do not operate any predictive AI systems. In fact, the wartime environment, mass displacement of the population, and exceptional circumstances have significantly hindered the adoption of new technologies in policing. Apart from Cassandra and the aforementioned judicial tools, AI in Ukrainian criminal justice thus remains at an early stage of development. This situation creates a certain ambiguity: on the one hand, Ukraine can observe and learn from the experiences (and mistakes) of other countries before implementing AI domestically; on the other, the risk remains that Ukrainian authorities may import solutions developed abroad—whether by private suppliers or international partners—that fail to reflect domestic legal specificities if internal regulatory frameworks do not keep pace.[31]

Overall, Ukraine's legal framework for AI in criminal proceedings continues to evolve, combining general legislation, ethical codes, and the influence of European standards that are currently shaping practice. In the absence of a specific Ukrainian statute regulating AI, oversight is grounded in existing legal principles. The Constitution of Ukraine guarantees the right to a fair trial and respect for human dignity—providing a fundamental safeguard against any uncontrolled automated decision-making that might compromise these rights. The national Data Protection Act, which largely mirrors the principles of the European GDPR, likewise discourages fully automated profiling in criminal matters without consent or explicit legal authorisation. Moreover, as a member of the Council of Europe, Ukraine is politically and legally bound to observe instruments such as the CEPEJ Ethical Charter (not legally binding but normatively persuasive) and the jurisprudence of the European Court of Human Rights. For example, if an AI tool were to be deployed in a way that undermined the

30 Bernaziuk (n 24).

31 Oleksandr Halahan and others, 'Digitalization of the Criminal Process: Is for the Better?' (2023) 38 IDP Revista de Internet, Derecho y Política 1. doi:10.7238/idp.v0i38.408495.

fairness of a trial, such use could amount to a violation of Article 6 of the European Convention on Human Rights.[32]

The Code of Judicial Ethics (Article 16) effectively establishes a strict limitation: judges may not delegate their decision-making authority to AI, particularly with regard to the evaluation of evidence or the rendering of judgments. This rule functions as a safeguard in the adjudicatory phase of proceedings. By contrast, the use of AI at earlier stages—such as by the police or within correctional institutions—is being closely monitored. Observers emphasise that, absent reliable safeguards, AI should remain a purely supportive instrument in criminal proceedings.[33]

The *Cassandra* example encapsulates many of these concerns. Issues of bias (does the algorithm overestimate risks for certain minorities?), transparency (can defendants challenge the manner in which their risk is calculated?), and contestability (what remedies are available to individuals who believe a decision influenced by AI was unjust?) are central to the Ukrainian debate and echo broader European discussions about algorithmic justice. In response, there have been growing calls for Parliament to adopt legislation specifically addressing AI in the judiciary.[34] Such a framework could, for instance, clarify the evidentiary status of AI-generated materials (for example, whether they constitute expert evidence or merely recommendations), establish transparency obligations (including disclosure to defendants), and establish oversight mechanisms such as certification or judicial council approval.

As of 2025, Ukraine stands at a crossroads. While it broadly endorses European principles and has introduced ethical restrictions, concrete legislative measures and systematic oversight mechanisms have yet to catch up with the few AI tools already in use.

6 SAFEGUARDING RIGHTS, EVIDENCE, AND FAIR TRIAL STANDARDS

Having outlined the introduction and governance of AI in Germany and Ukraine, the discussion now turns to several cross-cutting issues that are fundamental to ensuring that the deployment of AI in criminal justice does not undermine fundamental rights or the integrity of legal processes. These include algorithmic bias and explainability, the

32 Tetyana Antsupova and Sergii Koziakov, 'News Digest No 2 on Ukraine Judiciary: (research project The Dynamicsof the Judiciary in Ukraine in the Context of the Rule of Law and the EU Accession Aspirations, September 15 – October 15, 2024)' (*Bingham Centre for the Rule of Law*, 19 August 2025) <https://binghamcentre.biicl.org/newsitems/185/news-digest-no-12-on-ukraine-judiciary> accessed 21 August 2025.

33 Oksana Kaplina and others, 'Application of Artificial Intelligence Systems in Criminal Procedure: Key Areas, Basic Legal Principles and Problems of Correlation with Fundamental Human Rights' (2023) 6(3) Access to Justice in Eastern Europe 147. doi:10.33327/AJEE-18-6.3-a000314

34 Zhukevych and others (n 25).

implications of predictive policing for the presumption of innocence and proportionality, and the challenges posed by evidence generated or processed by AI.

The risk of algorithmic bias constitutes a central concern in the use of AI within criminal justice. Both Germany and Ukraine remain mindful of cautionary examples from other jurisdictions, particularly the United States, where investigations into the COMPAS risk-assessment tool revealed that Black defendants were disproportionately classified as high-risk compared to white defendants. Such findings underscore the danger that AI systems may unintentionally entrench or even exacerbate historical biases embedded in data. If left unaddressed, this risk could undermine the principle of equality before the law.[35]

Accordingly, in Europe, explainability and contestability are regarded as non-negotiable safeguards. The EU Artificial Intelligence Act effectively requires that high-risk AI systems—such as those deployed in the justice sector—be designed with explainability in mind, meaning their outputs must be interpretable and understandable to humans and subject to continuous human oversight.[36] Similarly, the Council of Europe standards demand transparency concerning the functioning of AI instruments.[37]

From a practical perspective, explainability is essential to ensure that judges, lawyers, and defendants can understand the recommendations produced by AI. For instance, if Ukraine's Cassandra tool classifies a probationer as "high risk," the probation officer and the court must be able to determine whether this assessment was driven by the individual's prior convictions, employment status, or some other factor—and this must be communicated in accessible language rather than as a mere numerical score. Contestability goes hand in hand with explainability: the affected person (or their counsel) must have the opportunity to challenge the algorithm's suggestion. At present, however, neither Germany nor Ukraine has clear procedural rules for contesting AI-generated information in court.[38] One can imagine, for example, a defence lawyer filing a motion to disclose the parameters of the algorithm or requesting an expert to scrutinise its methodology. In practice, a robust system might in the future require that any use of AI in criminal proceedings be disclosed to all parties, with input and output records preserved for potential review. Such measures would safeguard the right to a fair trial in the age of AI: ensuring that no decision rests on blind reliance on a machine, and that every statement influencing a case can be examined and contested.[39]

Predictive policing tools, such as those currently tested in Germany, raise unique legal and ethical challenges. One concern relates to proportionality: the deployment of resource-

35 Julia Angwin and others, 'Machine Bias' (*ProPublica*, 23 May 2016) <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> accessed 10 August 2025.

36 Regulation (EU) 2024/1689 (n 3).

37 Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (n 6).

38 Fair Trials (n 27).

39 Jacobson (n 28).

intensive, privacy-invasive surveillance measures on the basis of algorithmic forecasts must demonstrably yield tangible security benefits to justify their impact on citizens' privacy and freedom of movement.[40] German experiments with predictive policing to date have shown, at best, modest benefits—such as slight reductions in theft in certain areas—casting doubt on the compatibility of such systems with the constitutional requirements of necessity and proportionality. A further issue concerns the presumption of innocence and the closely related principle that law enforcement authorities must not pursue individuals without concrete suspicion. Is it sufficient, for instance, for an algorithm to identify a "high-risk area" or a "high-risk person" on the basis of data correlations to initiate a police investigation? German legal doctrine tends towards a negative answer: algorithmic risk profiles alone cannot substitute for individualised suspicion.[41] As a result, tools such as PRECOBS have been deliberately confined to supplementing, rather than replacing, traditional investigative methods, and police are generally instructed not to arrest or prosecute individuals solely because an algorithm flagged them.[42]

To safeguard rights in this context, experts recommend multi-layered oversight. In Germany, any sustained use of predictive policing algorithms must be accompanied by regular independent evaluations of their effectiveness and bias. A Data Protection Impact Assessment (DPIA) is, or soon will be, mandatory under the Artificial Intelligence Act and the GDPR whenever such tools are used extensively to process personal data, thereby ensuring thorough scrutiny of their privacy implications. Furthermore, proposals have been made to involve local legislatures or city councils in authorising and reviewing predictive policing pilot projects, in order to provide democratic oversight. Ukraine, by contrast, has not yet introduced predictive policing, giving it the advantage of learning before taking that step. Should Ukrainian authorities consider deploying such instruments (perhaps to relieve overburdened police forces), they could develop a legal framework in advance—for example, requiring judicial authorisation for certain types of AI-based surveillance, or imposing a moratorium on predictive systems until thorough public consultations and pilot studies have been completed.

In both countries, it is essential that crime prediction algorithms, if used, do not become covert tools of profiling or arbitrary enforcement. Predictive alerts should ideally be treated as leads requiring traditional verification, with human judgment firmly at the forefront of law enforcement—consistent with both Germany's and Ukraine's emphasis on human oversight in criminal justice.

With the growing prevalence of AI tools, courts will increasingly encounter evidence generated, influenced, or processed by algorithms. This may include, for example, AI-based translations of foreign-language materials, transcripts of audio recordings produced by

40 Vepřek and others (n 14).
41 Spell (n 15).
42 PRECOBS (n 13).

speech-recognition software, or AI-enhanced images—such as sharpened surveillance footage. In the near future, such issues may extend even further, encompassing the need to detect "deepfake" audio or video fabrications. German and Ukrainian courts alike will need to adapt evidentiary rules to meet these scenarios.

A pressing concern is authentication: how can a court be confident that a digital exhibit is genuine and has not been altered by AI? And if it has been altered—even for ostensibly legitimate purposes such as enhancing clarity—how should this affect its admissibility or evidentiary weight? At present, Germany has no official protocol governing AI-modified evidence, though these questions are increasingly acknowledged. Scholars caution that judges must remain vigilant regarding the provenance of digital materials and may require expert assistance in verifying authenticity. For instance, if prosecutors present a video that has been enhanced or edited by software, the defence should be entitled to review the original and to question an expert about the reliability of the software used.

Another issue concerns the interpretation of AI-generated results submitted as evidence. Suppose an algorithm scans a hard drive and flags 1,000 out of 50,000 images as likely illegal. Should these flagged files automatically be accepted as proof of an offence? Certainly not: human investigators must review each image individually, and in court the mere fact that the algorithm pre-selected them may be irrelevant—or even prejudicial—if not handled with caution. In Ukraine, where digital forensics is still in its infancy, reliance on AI tools supplied by external partners (for example, Interpol or Europol instruments in combating cybercrime) means that Ukrainian courts sometimes receive analytical reports not produced domestically. Judges, therefore, require guidance from the Ministry of Internal Affairs on how to handle such evidence.

Both countries would benefit from publishing practical guidelines or handbooks for judges on AI-related evidence. Such guidance might include: obligations to disclose any use of AI in evidence processing; guarantees that the defence has an opportunity to examine the algorithm or its outputs; and instructions that any uncertainty arising from AI processing should be resolved in favour of the defendant, consistent with the principle of in dubio pro reo. As a proactive measure, the Federal Ministry of Justice in Germany could convene a working group to examine these questions and propose necessary amendments to evidentiary rules. Ukraine, which often looks to Germany and other European states for models of best practice, could adopt similar measures.

In all cases, safeguarding the integrity of fact-finding remains paramount: AI must not become a "black box" through which evidence is introduced into court without being subjected to the same adversarial scrutiny as evidence produced by human means.

7 CONCLUSION

The introduction of artificial intelligence into criminal justice is a delicate undertaking, as illustrated by the approaches taken in Germany and Ukraine. Both countries are introducing AI incrementally—Germany through carefully controlled initiatives emphasising infrastructure and administrative efficiency, and Ukraine through targeted tools combined with explicit ethical restrictions. In neither jurisdiction is there any rush to delegate core judicial powers to algorithms. On the contrary, there is a clear commitment to maintaining a "human-in-the-loop" approach, consistent with European values and international recommendations. This reflects the shared understanding that, regardless of their sophistication, algorithmic systems lack the moral judgment, accountability, and contextual reasoning that human judges and officials bring to the judicial process.

The challenges, however, remain substantial. Germany's experience shows that even seemingly narrow applications such as predictive policing can raise complex questions of oversight and rights protection, while Ukraine's early adoption of a risk-assessment tool highlights the difficulties of ensuring transparency and public trust under conditions of limited resources. Both jurisdictions must translate high-level principles into practice: Germany by carrying out audits and evaluations, potentially restricting or recalibrating tools that fail to meet legal standards, and Ukraine by reinforcing its framework through legislation and independent oversight to prevent gaps as technologies expand. The entry into force of the EU Artificial Intelligence Act is likely to accelerate these developments by establishing a regulatory structure that underpins many of the issues discussed in this article.

What emerges is that AI-based technologies carry both promise and peril, particularly with respect to the protection of human rights. Their introduction requires a precise and optimal balance.

The comparative analysis of Germany and Ukraine underscores that the responsible integration of AI in criminal justice requires a principled, thematic approach. The following recommendations, grouped under core themes, provide a practical framework for policymakers and judicial actors:

1. Human-Centred Justice. AI may support information processing, but must never substitute for judicial reasoning. Decisions affecting rights and liberties must remain attributable to human judges or officials, ensuring dignity, accountability, and individualised reasoning.

2. Transparency and Explainability. The deployment and operation of AI tools should be publicly documented, with clear model cards or equivalent documentation. In court proceedings, parties must receive case-specific explanations of AI outputs to allow meaningful contestation. Black-box systems are incompatible with the criminal justice system.

3. Accountability and Oversight. Every AI tool should have a designated authority responsible for its operation, subject to regular internal and external audits. Independent oversight bodies—such as data protection authorities or ethics councils—must have powers to investigate and enforce recommendations. Logs of AI processes should be preserved for potential review.

4. Data Protection and Privacy. Given the sensitivity of personal and biometric data in criminal justice, systems must comply with the highest standards of data minimisation, anonymisation, and cybersecurity. Secondary use of judicial data, particularly for commercial training purposes, should be strictly prohibited unless expressly authorised.

5. Scope and Proportionality. AI programs must be legally authorised for a defined purpose and not repurposed without renewed scrutiny. Predictive policing or surveillance tools should be deployed only as pilot projects, subject to proportionality tests and democratic authorisation.

Ultimately, the comparative analysis of Germany and Ukraine underscores that "intelligent" criminal justice is not about replacing human judgment with machine calculations but about finding safe and effective ways to assist judges, lawyers, and law enforcement. Thoughtfully implemented, AI can indeed enhance efficiency by filtering data, reducing delays, and promoting consistency. Implemented carelessly, however, it risks undermining rights and eroding trust in the justice system. The way forward, as both countries demonstrate, lies in a balanced approach: embracing innovation while embedding every step in oversight, transparency, and an unwavering commitment to the values of justice.

This perspective resonates with broader global debates: UNESCO's ethical frameworks and interdisciplinary AI governance research converge on the same imperative—embedding transparency, fairness, and human accountability at the heart of judicial innovation.[43] This alignment is further reinforced by the OECD/G20 AI Principles[44] and interdisciplinary analyses from political science and philosophy, which emphasise the broader societal risks of opacity, accountability gaps, and the erosion of trust.[45] These perspectives extend the comparative findings of this article into the wider global conversation on responsible AI governance. In an era of rapid technological change, preserving a human-centred judiciary and respect for rights is both the most significant challenge and the ultimate goal.

43 UNESCO (n 7); Floridi and Cowls (n 9).

44 G20 (n 10); OECD (n 10).

45 Joanna J Bryson, 'The Past Decade and Future of AI's Impact on Society' in *Towards a New Enlightenment?: A Transcendent Decade* (Turner 2019); Corinne Cath, 'Governing Artificial Intelligence: Ethical, Legal and Technical Opportunities and Challenges' (2018) 376(2133) Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 20180080. doi:10.1098/rsta.2018.0080.

REFERENCES

1. Angwin J and others, 'Machine Bias' (*ProPublica*, 23 May 2016) <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> accessed 10 August 2025
2. Antsupova T and Koziakov S, 'News Digest No 2 on Ukraine Judiciary: (research project The Dynamicsof the Judiciary in Ukraine in the Context of the Rule of Law and the EU Accession Aspirations, September 15 – October 15, 2024)' (*Bingham Centre for the Rule of Law*, 19 August 2025) <https://binghamcentre.biicl.org/newsitems/185/news-digest-no-12-on-ukraine-judiciary> accessed 10 August 2025
3. Bernaziuk I, 'Artificial Intelligence and the Judicial System of Ukraine: Results of Cooperation in the Past Year: Presentation' (JAR-Association Conference "Judicial Systems in Transition: Reforms, Innovations and Justice", JAR-Association, OMIJ, Faculty of Law University of Limoges, 11 June 2025) <https://court.gov.ua/storage/portal/supreme/prezentacii_2025/AI_Ukraine_bernaziuk.pdf> accessed 10 August 2025
4. Bryson JJ, 'The Past Decade and Future of AI's Impact on Society' in *Towards a New Enlightenment?: A Transcendent Decade* (Turner 2019)
5. Cath C, 'Governing Artificial Intelligence: Ethical, Legal and Technical Opportunities and Challenges' (2018) 376(2133) Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 20180080. doi:10.1098/rsta.2018.0080
6. Floridi L and Cowls J, 'A Unified Framework of Five Principles for AI in Society' (2019) 1(1) Harvard Data Science Review 1. doi:10.1162/99608f92.8cd550d1
7. Halahan O and others, 'Digitalization of the Criminal Process: Is for the Better?' (2023) 38 IDP Revista de Internet, Derecho y Política 1. doi:10.7238/idp.v0i38.408495
8. Izarova I and others, 'Advancing Sustainable Justice Through AI-Based Case Law Analysis: Ukrainian Experience' (2024) *7*(1) *Access to Justice in Eastern Europe* 127. doi:10.33327/AJEE-18-7.1-a000123
9. Jacobson J, 'Algorithmic Risk Assessment Tools in Criminal Proceedings: An Analysis in Light of Articles 6 and 14 of the European Convention on Human Rights' (Master's thesis, Uppsala University 2022).
10. Kaplina O and others, 'Application of Artificial Intelligence Systems in Criminal Procedure: Key Areas, Basic Legal Principles and Problems of Correlation with Fundamental Human Rights' (2023) 6(3) Access to Justice in Eastern Europe 147. doi:10.33327/AJEE-18-6.3-a000314
11. Kiesow Cortez E and Maslej N, 'Adjudication of Artificial Intelligence and Automated Decision-Making Cases in Europe and the USA' (2023) 14(3) European Journal of Risk Regulation 457. doi:10.1017/err.2023.61

12. Mittelstadt B, 'Principles Alone Cannot Guarantee Ethical AI' (2019) 1 Nature Machine Intelligence 501. doi:10.1038/s42256-019-0114-4

13. Moskvytyn Yu and Limante A, 'Integrating Artificial Intelligence in Ukraine's Courts: State of Play and Future Prospect' (*Verfassungsblog*, 12 December 2024) <https://verfassungsblog.de/ai-ukraine-judiciary/> accessed 10 August 2025

14. Shi C, Sourdin T and Li B, 'The Smart Court – A New Pathway to Justice in China?' (2021) 12(1) International Journal for Court Administration 4. doi:10.36745/ijca.367

15. Spell A, 'The Use of Predictive Policing in German Law Enforcement: A Discourse Analysis' (Bachelor's thesis, University of Twente 2023)

16. Sprenger J and Brodowski D, '"Predictive Policing", "Predictive justice", and the Use of AI in the Administration of Justice in Germany' [2023] e-Revue Internationale de Droit Pénal 117. doi:10.22028/D291-39980

17. Vepřek LH and others, 'Legitimising Predictive Policing in Germany' (2020) 2(3) Kriminologie 1. doi:10.18716/ojs/krimoj/2020.3.3

18. Zhukevych I and others, 'Analysis of Issues Related to the Legalization of Artificial Intelligence, Its Use in Legal Proceedings, Legal Consultation and Law Enforcement' (2024) 27 Legal Scientific Journal 17

AUTHORS INFORMATION

Lidiia Moskvych*
Dr Sc (Law), Professor, Associate Professor of the Department of Criminal Procedure, Faculty of Prosecutor's Office, Yaroslav Mudryi National Law University, Kharkiv, Ukraine
l.m.moskvych@nlu.edu.ua
https://orcid.org/0000-0001-7339-3982
Corresponding author, responsible for research methodology, writing and management.

Competing interests: No competing interests were reported.

Disclaimer: The author declares that her opinions and views expressed in this manuscript are free from any influence of any organization, including the Constitutional Court of Ukraine, despite the fact that she is a member of the Scientific Advisory Council of the Constitutional Court.

Iryna Borodina
Cand. of Science of Law (Equiv. Ph.D.), Associate Professor, Associate Professor of the Department of Criminal Procedure, Faculty of Prosecutor's Office, Yaroslav Mudryi National Law University, Kharkiv, Ukraine
i.v.borodina@nlu.edu.ua
https://orcid.org/0009-0008-7611-6575
Co-author, responsible for data collection and writing.

Competing interests: No competing interests were announced by the author.

Disclaimer: The author declares that her opinion and views expressed in this manuscript are free of any impact of any organizations.

Olga Ovsiannikova
Cand. of Science of Law (Equiv. Ph.D.), Associate Professor, Associate Professor of the Department of Criminal Procedure, Faculty of Prosecutor's Office, Yaroslav Mudryi National Law University, Kharkiv, Ukraine
o.o.ovsiannikova@nlu.edu.ua
https://orcid.org/0000-0001-7773-6487
Co-author, responsible for data collection and writing.

Competing interests: No competing interests were announced by the author.

Disclaimer: The author declares that her opinion and views expressed in this manuscript are free of any impact of any organizations.

RIGHTS AND PERMISSIONS

EDITORS

ABOUT THIS ARTICLE

Cite this article
Moskvych L, Borodina I and Ovsiannikova O, 'Artificial Intelligence in Criminal Justice in Germany and Ukraine: A Comparative Legal Study' (2025) 8(Spec) Access to Justice in Eastern Europe 210-32 <https://doi.org/10.33327/AJEE-18-8.S-a000155> P

DOI: https://doi.org/10.33327/AJEE-18-8.S-a000155

Keywords: *Artificial Intelligence; Criminal Justice; Predictive Policing; Risk Assessment; Explainability; Data Protection; Human Oversight.*

DETAILS FOR PUBLICATION

Date of submission: 03 Sep 2025
Date of acceptance: 01 Oct 2025
Date of Publication: 30 Dec 2025
Whether the manuscript was fast tracked? - No
Number of reviewer report submitted in first round: 2 reports
Number of revision rounds: 1 round with minor revisions

Technical tools were used in the editorial process:
Plagiarism checks - Turnitin from iThenticate https://www.turnitin.com/products/ithenticate/
Scholastica for Peer Review https://scholasticahq.com/law-reviews

АНОТАЦІЯ УКРАЇНСЬКОЮ МОВОЮ

Дослідницька стаття

ШТУЧНИЙ ІНТЕЛЕКТ У КРИМІНАЛЬНОМУ ПРАВОСУДДІ НІМЕЧЧИНИ ТА УКРАЇНИ: ПОРІВНЯЛЬНО-ПРАВОВЕ ДОСЛІДЖЕННЯ

Лідія Москвич, Ірина Бородіна та Ольга Овсяннікова

АНОТАЦІЯ

Вступ*. Штучний інтелект (ШІ) швидко переходить від допоміжних адміністративних інструментів до застосувань, які безпосередньо впливають на функціонування систем кримінального правосуддя. У Європі цей процес відбувається в межах обережного, правоорієнтованого підходу, покликаного поєднати інновації з дотриманням принципів незалежності суду, справедливості та верховенства права. Дослідження пропонує порівняльно-правовий аналіз впровадження ШІ у системи кримінального правосуддя Німеччини та України в контексті Акту ЄС про штучний інтелект (2024), стандартів*

Ради Європи та конституційних гарантій. Структурований і скоординований на федеральному рівні підхід Німеччини контрастує з більш точковою, але етично обмеженою моделлю України, що відображає відмінності інституційних спроможностей і правових традицій.

***Методи**. Робота ґрунтується на порівняльно-правовому підході, що поєднує функціональний та контекстуальний аналіз. Функціональний аспект виявляє, як Німеччина та Україна застосовують ШІ для вирішення подібних проблем — підвищення ефективності, прозорості та захисту прав. Контекстуальний аспект розглядає вплив конституційних, інституційних та соціально-політичних чинників, зокрема умов воєнного стану в Україні. Такий комбінований підхід дає змогу оцінювати подібності й відмінності не абстрактно, а в контексті європейської та національних правових культур.*

Аналіз базується на законодавстві, офіційних звітах, судовій практиці та наукових джерелах. Емпіричну базу становлять, зокрема, німецькі пілотні проєкти у сфері прогностичної поліцейської діяльності (PRECOBS, KLB-operativ), інструменти фільтрації слідчих даних і адміністративні ШІ-рішення в судах, а також український алгоритм оцінки ризиків у пробації Cassandra та ШІ-інструменти для правничих досліджень і перекладу. Досвід Сполучених Штатів щодо алгоритмічної оцінки ризиків використовується як застережливий орієнтир.

***Результати та висновки**. Дослідження показує, що обидві юрисдикції обмежують використання ШІ як заміни у процесі прийняття рішень. Водночас у Німеччині використання ШІ координується на федеральному рівні та обмежується допоміжними функціями — передусім у сфері адміністративної оптимізації й аналітичної підтримки. Натомість в Україні впровадження ШІ є більш вибірковим і підпорядкованим чітким етичним обмеженням, але ускладнюється нестачею прозорості та об'єктивними обмеженнями, зумовленими воєнним станом. Аналіз виявляє спільні проблеми, зокрема алгоритмічну упередженість, пояснюваність рішень, допустимість доказів і забезпечення гарантій справедливого судового розгляду, та формулює рекомендації з урахуванням національного контексту. Серед них — запровадження обов'язкових зовнішніх аудитів, законодавче закріплення процесуальних прав на оскарження даних, згенерованих ШІ, чіткі правила доказового використання таких даних, а також підвищення обізнаності суддів щодо технологій ШІ.*

Дослідження підкреслює, що сталий розвиток ШІ у кримінальному правосудді можливий лише за умови його допоміжного характеру, прозорості, можливості аудиту та збереження людського контролю, що є необхідним для відповідності європейським правовим стандартам і захисту основоположних прав.

***Ключові слова**. Штучний інтелект, кримінальне правосуддя, прогностичні алгоритми у поліцейській діяльності, оцінка ризиків, пояснюваність, захист даних, людський контроль.*

Access to Justice in Eastern Europe
ISSN 2663-0575 (Print)
ISSN 2663-0583 (Online)
http://ajee-journal.com

Research Article

ARTIFICIAL INTELLIGENCE IN CRIMINAL JUSTICE: BALANCING TECHNOLOGICAL INNOVATION AND PERSONAL DATA PROTECTION RIGHTS. A COMPARATIVE LEGAL STUDY BETWEEN THE EUROPEAN UNION AND VIETNAM

Phuong Anh Nguyen

DOI:
https://doi.org/10.33327/AJEE-18-8.S-r000161

Date of submission: 26 Oct 2025
Date of acceptance: 08 Dec 2025
Online First publication: 19 Dec 2025
Last Published: 30 Dec 2025

Disclaimer:

The author declares that their opinion and views expressed in this manuscript are free of any impact of any organizations.

ABSTRACT

Background: *The rapid development of Artificial Intelligence has profoundly transformed various aspects of social life, including the criminal justice system. In criminal proceedings, the collection and processing of biometric, behavioural, and emotional data may threaten the right to privacy, the presumption of innocence, and the right to a fair trial. This study examines the intersection between technological innovation and personal data protection in criminal justice through a comparative legal analysis of the European Union and Vietnam. By analysing the EU's GDPR, Law Enforcement Directive, and AI Act 2024 alongside Vietnam's legal framework, the paper identifies key areas of convergence, divergence, and regulatory gaps.*

Methods: *The study employs comparative legal analysis, combined with a human rights–based approach, to clarify the relationship between technological innovation and the right to personal data protection in criminal justice. The sources of reference include the legal frameworks of the European Union and Vietnam, case law, and reports from EU and United Nations bodies.*

Results and Conclusions: *The study aims to establish fundamental legal principles that balance technological innovation with the protection of the right to personal data in criminal justice—an approach that has received limited attention in Vietnam. Based on this foundation, it proposes legal reforms toward a framework of "human rights–oriented digital justice", ensuring that the digitalisation and application of AI in the justice system not only enhance operational efficiency but also strengthen the rule of law and protect people.*

1 INTRODUCTION

The rapid development of Artificial Intelligence (AI) is creating profound changes in criminal justice activities worldwide. In many countries, AI tools have been tested or deployed at various stages of criminal proceedings—such as investigation, prosecution, and adjudication — to improve the efficiency of legal proceeding activities, while minimising "unintentional" or "intentional" errors made by humans. These technologies assist competent authorities in making judicial decisions more quickly, transparently, and objectively. Such efforts not only reflect the trend of modernising the justice systems but also represent the broader digital revolution that is reshaping how the State exercises judicial power in the 21st century.[1] However, alongside the opportunity for innovation, AI technologies also expose an increasingly clear contradiction between two fundamental legal objectives: technological innovation in judicial activities and the protection of human rights, particularly the right to data protection. AI governance, therefore, cannot only remain at the level of ethical recommendations; it must be firmly grounded in the rule of law. AI should be guided to serve human development rather than merely optimise efficiency.[2] This reality underscores the urgent need to integrate human values, accountability, and transparency into the justice system's digital transformation.

The EU is a pioneer in establishing a legal framework that balances technological innovation and personal data protection. The Artificial Intelligence Act (AI Act) of 2024 is the first legal document in the world to comprehensively regulate AI based on a risk-classification approach, in which applications in the fields of justice and law enforcement are categorised

1 Angela Daly and others, 'Artificial Intelligence, Governance and Ethics: Global Perspectives' (Research Paper no 2019-15, The Chinese University of Hong Kong Faculty of Law 2019) doi:10.2139/ssrn.3414805.

2 Bernd Carsten Stahl and others, 'Artificial Intelligence for Human Flourishing – Beyond Principles for Machine Learning' (2021) 124 Journal of Business Research 374. doi:10.1016/j.jbusres.2020.11.030.

as "high-risk."[3] Alongside this, the Directive 2016/680/EU[4] (Law Enforcement Directive – LED) and the Regulation 2016/679/EU[5] (General Data Protection Regulation – GDPR) reaffirm the position of "the right to personal data protection" as a fundamental right, guaranteed in all data processing activities conducted for criminal proceedings.

In Vietnam, digital transformation in judicial activities has been institutionalised through a series of policies and legal documents. Resolution No. 27-NQ/TW of the Central Committee of the Communist Party of Vietnam[6] affirms the requirement to modernise judicial activities and apply new technologies while simultaneously respecting and protecting human rights. The 2015 Criminal Procedure Code for the first time recognises electronic data as a source of evidence (Articles 87 and 99),[7] paving the way for the application of information technology and AI in collecting, evaluating, and using evidence. Furthermore, Decree No. 13/2023/ND-CP on personal data protection[8] and the 2025 Law on Personal Data Protection[9] (effective from 1 January 2026) have established a unified legal foundation to protect individual rights in the digital space. However, as judicial authorities begin to utilise big data, biometrics, and AI-powered predictive analytics tools, a crucial question arises: how can new technologies be applied without "diminishing" the level of protection of fundamental human rights, especially the right to personal data protection?

3 Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying down Harmonised Rules on Artificial Intelligence and Amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act - AI Act) <http://data.europa.eu/eli/reg/2024/1689/oj> accessed 24 October 2025.

4 Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA (Law Enforcement Directive - LED) [2016] OJ L 119/89.

5 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation - GDPR) [2016] OJ L 119/1.

6 Resolution of the Central Committee of the Communist Party of Vietnam No 27-NQ/TW 'On Continuing to Build and Perfect the Socialist Rule of Law State of Vietnam in the New Period' (9 November 2022) [in Vietnamese] <https://thuvienphapluat.vn/van-ban/Bo-may-hanh-chinh/Nghi-quyet-27-NQ-TW-2022-tiep-tuc-xay-dung-Nha-nuoc-phap-quyen-xa-hoi-chu-nghia-giai-doan-moi-541092.aspx> accessed 24 October 2025.

7 Law of the Socialist Republic of Vietnam No 101/2015/QH13 'Criminal Procedure Code' (27 November 2015) <https://thuvienphapluat.vn/van-ban/Trach-nhiem-hinh-su/Bo-luat-to-tung-hinh-su-2015-296884.aspx> accessed 24 October 2025.

8 Decree of the Socialist Republic of Vietnam No 13/2023/ND-CP 'Protection of Personal Data' (17 April 2023) <https://thuvienphapluat.vn/van-ban/Cong-nghe-thong-tin/Nghi-dinh-13-2023-ND-CP-bao-ve-du-lieu-ca-nhan-465185.aspx> accessed 24 October 2025.

9 Law of the Socialist Republic of Vietnam No 91/2025/QH15 'Personal Data Protection' (26 June 2025) <https://thuvienphapluat.vn/van-ban/Bo-may-hanh-chinh/Luat-Bao-ve-du-lieu-ca-nhan-2025-so-91-2025-QH15-625628.aspx> accessed 24 October 2025.

Based on the practice of the EU and Vietnam, this study focuses on clarifying: (i) the theoretical framework and legal principles governing the relationship between AI and personal data in criminal proceedings; (ii) the EU's experience and balancing mechanism as reflected in typical legal documents and case law; and (iii) suggestion for improving Vietnamese law to ensure that technological innovation is aligned with the protection of the right to personal data. This study not only systematises Vietnam's legal framework amid the digital transformation of justice but also proposes a balanced model in which technological innovation and personal data protection reinforce one another to build a modern, humane, and trustworthy criminal justice system.

2 METHODOLOGY

The article adopts the doctrinal legal research method combined with comparative legal analysis, grounded in a human rights–based approach. The EU is one of the most successful regions in the world in establishing both a comprehensive legal framework and effective mechanisms for the protection of human rights in general and of personal data in particular. For this reason, the author has chosen the EU model as a good-practice example to analyse and draw relevant lessons for improving Vietnam's legal framework in this field. To achieve this objective, the research sources include legal documents of the EU, including the GDPR, Law Enforcement Directive (LED), AI Act, and Convention 108+, and related legal documents of Vietnam, along with the case law of the European Court of Human Rights (ECtHR) and the Court of Justice of the European Union (CJEU). This methodology aims to explain, compare, and generalise key legal principles, especially the principles of lawfulness, necessity, and proportionality, as well as accountability in the application of AI in criminal justice. The article does not use empirical data; rather, it focuses on normative analysis and academic arguments, thereby providing directions for improving the law and policy implications tailored to Vietnam's context.

3 CONCEPT FRAMEWORK

3.1. The Definition of "Artificial Intelligence in Criminal Justice"

In legal science, there is no absolute unified definition of AI. However, within the European legal framework, the commonly adopted understanding of AI is a system capable of "reasoning, learning, and generating predictions, recommendations, or decisions that may affect real or virtual environments" based on the analysis of pre-programmed input data.[10]

10 EU-LISA and Eurojust, *Artificial Intelligence Supporting Cross-Border Cooperation in Criminal Justice* (Publications Office of the EU 2022). doi:10.2857/364146.

In the context of criminal justice, AI refers to the use of fully or semi-automated digital systems to assist humans with procedural activities such as collecting, analysing, and evaluating evidence; managing case files; forecasting the risk of recidivism; or monitoring the execution of sentences. The goal of applying AI is to enhance the efficiency of crime detection and handling while reducing the workload of judicial and prosecutorial authorities.

However, as these activities directly affect individual freedoms and the right to a fair trial, the application of AI in the criminal justice system must always be subject to strict legal limits and human oversight. Since AI systems used in law enforcement and criminal justice are classified as "high-risk", they are subject to additional obligations, such as assessing conformity, record-keeping, ensuring transparency and explainability of outputs, and "human oversight," which means humans must always retain ultimate control.[11]

3.2. The Definition of "Technological Innovation in Criminal Justice"

The concept of technological innovation in criminal justice not only refers to the adoption of new technologies but also implies a fundamental transformation in the process of exercising judicial authority. From an academic perspective, technological innovation in this field can be identified through three key legal criteria:

First, the degree of automation and its impact on procedural decisions. The more capable the system is of making decisive conclusions or suggestions (for example: analysing DNA samples, recognising a suspect's facial features, or recommending sentencing based on precedent data), the higher the legal requirements regarding transparency, explainability, and verifiability.[12]

Second, the type of data being processed. In the field of criminal justice, most data is classified as sensitive, especially biometric, health, and personal data, as well as information concerning political or religious beliefs. According to the EU Directive 2016/680 on data protection in law enforcement activities (the Law Enforcement Directive – LED), the processing of such data is lawful only if there is a clear legal basis that complies with the principles of necessity and proportionality.[13]

Third, the legal consequences for human rights. Every application of technology in criminal proceedings must be assessed through the lens of the right to a fair trial, the right to privacy, and the right to personal data protection. The use of surveillance or predictive analysis tools without adequate safeguards and legal boundaries may pose risks of human rights violations. The European Court of Human Rights, in *S. and Marper v. the United Kingdom* (2008), affirmed that the indefinite and indiscriminate retention of biometric

11 Regulation (EU) 2024/1689 (n 3).

12 ibid

13 Directive (EU) 2016/680 (n 4).

data from individuals who have not been convicted of a crime is a disproportionate interference with the right to respect for private life under Article 8 of the European Convention on Human Rights.[14]

3.3. The Definition of "Right to Personal Data Protection in Criminal Justice"

The right to personal data protection has a solid foundation in international human rights law. According to provisions in Article 12 of the Universal Declaration of Human Rights (1948),[15] Article 17 of the International Covenant on Civil and Political Rights (1966),[16] and Article 8 of the European Convention on Human Rights (1950),[17] it can be seen that the principles of lawfulness, necessity, and proportionality must govern the collection, storage, and use of personal information. In particular, the development of Convention 108+ (2018) has held that the "right to personal data protection" is an independent right, closely linked to the "right to privacy".[18] This represents a new generation of fundamental human rights designed to address the challenges posed by the era of AI and big data.

In the criminal field, Directive 2016/680/EU establishes a distinct protection mechanism for the processing of personal data by competent authorities for investigation, prosecution, and adjudication. It allows for certain restrictions of individual rights, such as the right to access or object, but only to the extent necessary to avoid obstructing judicial activities. At the same time, it requires the establishment of independent monitoring mechanisms and guarantees effective rights to appeal.[19]

In Vietnam, this right has recently been institutionalised by Decree No. 13/2023/ND-CP and the 2025 Law on Personal Data Protection. These documents set forth seven fundamental principles governing the processing of personal data, including "collecting data only as necessary for legitimate purposes", "ensuring the accuracy and timeliness of data", and "retaining data only for the necessary period".[20] The 2025 Law on Personal Data Protection further reaffirms the fundamental rights of data subjects and

14 *S and Marper v the United Kingdom* Apps nos 30562/04, 30566/04 (ECtHR, 4 December 2008) <https://hudoc.echr.coe.int/fre?i=001-90051> accessed 24 October 2025.

15 Universal Declaration of Human Rights (adopted 10 December 1948 UNGA Res 217A) <https://www.un.org/en/about-us/universal-declaration-of-human-rights> accessed 24 October 2025.

16 International Covenant on Civil and Political Rights (adopted 16 December 1966 UNGA Res 2200A(XXI)) 999 UNTS 171.

17 Council of Europe, *European Convention on Human Rights (Convention for the Protection of Human Rights and Fundamental Freedoms, as amended by Protocols)* (ECHR 2013).

18 Council of Europe, *Convention 108+ Convention for the Protection of Individuals with Regard to the Processing of Personal Data* (Council of Europe 2018) <https://www.coe.int/en/web/data-protection/convention108-and-protocol> accessed 24 October 2025.

19 Directive (EU) 2016/680 (n 4).

20 Decree of the Socialist Republic of Vietnam No 13/2023/ND-CP (n 8); Law of the Socialist Republic of Vietnam No 91/2025/QH15 (n 9).

establishes a mandatory Data Protection Impact Assessment (DPIA) mechanism for data processing activities that pose a risk of infringing on privacy rights, particularly in the fields of justice and law enforcement.

When AI is integrated into this process, from facial recognition and behavioural analysis to risk prediction, personal data becomes not only a tool for investigation but also central to the right to a fair trial and the principle of the presumption of innocence.[21] The application of AI in criminal justice must therefore be framed within the context of human rights, ensuring that all technological innovations adhere to the principles of lawfulness, necessity, proportionality, human supervision, and accountability, to prevent the automation of judicial power.[22] In this context, the Fundamental Rights Impact Assessment (FRIA) serves as a core legal instrument in AI governance, helping to balance technological innovation and the protection of personal data rights, particularly in investigation and adjudication activities.[23] This demonstrates that a human rights–based model of criminal justice is the most appropriate approach to harmonise technological innovation with the safeguarding of personal data protection.

Therefore, the right to personal data protection in criminal justice is not only an extension of the right to privacy in the digital environment, but also a legal instrument to ensure control over state power when applying AI in investigation, prosecution, and adjudication. This right serves both preventive and protective functions, aiming to maintain a balance between technological efficiency and human rights protection. On that basis, it is essential to establish a set of fundamental legal principles to guide all activities in designing, implementing, and monitoring judicial technologies. This helps ensure that innovation occurs within the rule of law and in conformity with international human rights standards.

4 KEY PRINCIPLES ENSURING A BALANCE BETWEEN TECHNOLOGICAL INNOVATION AND THE RIGHT TO PERSONAL DATA PROTECTION

4.1. The Principles of Lawfulness, Necessity, and Proportionality

The principles of lawfulness, necessity, and proportionality were developed and refined through the case law of the European Court of Human Rights (ECtHR) in its interpretation

21 Sandra Wachter and Brent Mittelstadt, 'A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI' (2019) 2 Columbia Business Law Review 494.

22 Alessandro Mantelero, 'AI and Big Data: A Blueprint for a Human Rights, Social and Ethical Impact Assessment' (2018) 34(4) Computer Law & Security Review 754. doi:10.1016/j.clsr.2018.05.017.

23 Alessandro Mantelero, 'The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, Legal Obligations and Key Elements for a Model Template' (2024) 54 Computer Law & Security Review 106020. doi:10.1016/j.clsr.2024.106020.

of Article 8 of the European Convention on Human Rights (ECHR). A series of cases, from *S. and Marper v. the United Kingdom*,[24] *Uzun v. Germany,*[25] *Roman Zakharov v. Russia*[26] *to Big Brother Watch and Others v. the United Kingdom*[27] has established a three-tier standard in interpreting Article 8 ECHR, consisting of: (i) lawfulness – any interference must have a clear and foreseeable legal basis; (ii) necessity – it must only be applied in a democratic society; and (iii) proportionality – the degree of interference must not exceed what is required to protect the public interests. Among these cases, *S. and Marper v. the United Kingdom* represents a pivotal and foundational case, in which the ECtHR affirmed that the indefinite retention of biometric data of individuals who have not been convicted is an "unjustified interference with the right to privacy", thereby violating Article 8 of the European Convention on Human Rights.[28]

In EU laws, the principle of proportionality has been elevated to a constitutional legal standard and has been institutionalised in legal documents governing emerging technologies, most notably the 2024 AI Act. According to Article 6 and Annexe III of this document, AI systems used in law enforcement, investigation, prosecution, adjudication, and the execution of sentences are classified as high-risk.[29] To be deployed, such systems must comply with a set of mandatory legal obligations, including ex ante risk assessment, ensuring transparency, explainability, and human supervision, as well as the post-market monitoring mechanisms to control the entire lifecycle of the system.

In Vietnam, although the principles of lawfulness, necessity, and proportionality have not yet been explicitly defined in the field of AI, the key elements of these principles are indirectly reflected in Decree No. 13/2023/ND-CP and the 2025 Law on Personal Data Protection. According to Article 3 of Decree No. 13/2023/ND-CP, the processing of personal data must adhere to the principles of "only collecting data within the scope necessary for legitimate purposes", ensuring that data are "up to date," and "only retaining data for the necessary period".[30] In particular, the 2025 Law on Personal Data Protection adds the obligation to conduct a Data Protection Impact Assessment (DPIA) for data processing and transfer activities, including the application of AI in investigation, prosecution, and adjudication.[31] Although Vietnam has not yet adopted a Fundamental Rights Impact Assessment (FRIA) mechanism like the EU's, the DPIA represents an important step toward integration and toward balancing technological innovation and human rights.

24 *S and Marper* (n 14).

25 *Uzun v Germany* App no 35623/05 (ECtHR, 2 September 2010) <https://hudoc.echr.coe.int/eng?i=001-100293> accessed 24 October 2025.

26 *Roman Zakharov v Russia* App no 47143/06 (ECtHR, 4 December 2015) <https://hudoc.echr.coe.int/fre?i=001-159324> accessed 24 October 2025.

27 *Big Brother Watch and Others v the United Kingdom* Apps nos 58170/13, 62322/14, 24960/15 (ECtHR, 25 May 2021) <https://hudoc.echr.coe.int/eng?i=001-210077> accessed 24 October 2025.

28 *S and Marper* (n 14).

29 Regulation (EU) 2024/1689 (n 3).

30 Decree of the Socialist Republic of Vietnam No 13/2023/ND-CP (n 8).

31 Law of the Socialist Republic of Vietnam No 91/2025/QH15 (n 9).

4.2. The Principle of Accountability and Impact Assessment

According to Article 5(2) of the General Data Protection Regulation 2016/679 (GDPR), data controllers are responsible for, and must be able to demonstrate, compliance with the data protection principles.[32] This responsibility is operationalised through the Data Protection Impact Assessment (DPIA), a preventive tool designed to identify and mitigate risks before technology is applied. The obligation to conduct a DPIA is stipulated in Article 35 of the GDPR and Article 27 of Directive 2016/680 (LED), requiring an assessment of any data processing activities that pose a high risk, including the use of new technologies or large-scale monitoring.[33]

Accountability and impact assessment mechanisms are ethical and legal tools that enable the state and public authorities to balance technological innovation with the protection of human rights, thereby establishing a Human Rights, Ethical, and Social Impact Assessment (HRESIA) framework for high-risk AI systems.[34] In the context of criminal proceedings, these tools are particularly essential for controlling the transparency, accuracy, and bias of AI systems used in investigation, adjudication, or risk assessment. In the case of *State v. Loomis,* the defendant was sentenced based on the results of the COMPAS software but was unable to challenge or verify the algorithm, providing a clear example of the risks of algorithmic justice.[35] In Europe, in *NJCM v. the Netherlands*, the Hague Court also held that the government's use of a risk prediction system without a DPIA and an independent accountability mechanism violated Article 8 of the ECHR.[36] The Court affirmed that a human rights impact assessment procedure must govern all AI applications in the public sector.

Vietnamese law has adopted a similar approach. Decree No. 13/2023/ND-CP establishes an obligation to conduct impact assessments for high-risk data processing activities, particularly in the areas of security, justice, and information technology.[37] Subsequently, the 2025 Law on Personal Data Protection sets out mechanisms for reporting, independent monitoring, and the accountability of state agencies in managing personal data.[38] This represents an important step toward establishing a risk control mechanism for data in criminal proceedings. However, Vietnam is currently only at the internal assessment stage and lacks an independent, transparent accountability mechanism similar to the EU's DPIA-FRIA model.

32 Regulation (EU) 2016/679 (n 5).

33 ibid; Directive (EU) 2016/680 (n 4).

34 Mantelero (n 22; 23).

35 *State v Loomis* 371 Wis.2d 235, 881 N.W.2d 749 [2016] Wisconsin Supreme Court <https://case-law.vlex.com/vid/state-v-loomis-no-888404547> accessed 24 October 2025.

36 *NJCM c The Netherlands C-09-550982-HA ZA 18-388* [2020] Rechtbank Den Haag ECLI:NL:RBDHA:2020:1878 <https://www.escr-net.org/wp-content/uploads/2020/09/ecli_nl_rbdha_2020_1878.pdf> accessed 24 October 2025.

37 Decree of the Socialist Republic of Vietnam No 13/2023/ND-CP (n 8).

38 Law of the Socialist Republic of Vietnam No 91/2025/QH15 (n 9).

4.3. The Principle of Transparency and Human Supervision

According to Article 14 and Annexe III of the 2024 AI Act, high-risk AI systems, including tools used in law enforcement and criminal proceedings, must be equipped with "human supervision", meaning that humans can supervise, intervene, or suspend the system's processing when necessary.[39] This obligation is not only technical but also reflects the principle of the rule of law: humans remain the ultimate decision-makers in the judicial process, thereby maintaining the legitimacy of state authority in the digital era.

The report *Human Rights Due Diligence for Digital Technology Use* published by the United Nations Office of the High Commissioner for Human Rights (OHCHR) emphasises that all government agencies employing digital or AI systems must conduct human rights impact assessments, ensure transparency regarding the purpose and operation mechanisms, and maintain human supervision throughout the technology's lifecycle, especially in areas that may affect personal freedoms, dignity, and the right to access to justice.[40] Similarly, both the OECD AI Principles[41] (issued in 2019 and updated in 2024) and the Council of Europe's Framework Convention on Artificial Intelligence, Human Rights, Democracy, and the Rule of Law[42] (2024) require countries to ensure that humans remain the ultimate responsible party. Humans should be able to intervene, suspend, or disable AI systems when there is a risk of human rights violations. These documents reflect a global shift from "autonomous AI" to "AI for people", aiming to maintain transparency, accountability, and the rule of law in all areas, particularly in criminal justice, where the impacts of technological decisions can directly affect individuals' fates and personal freedoms.

In Vietnam, although there has not been a legal document directly governing the use of AI in legal proceedings, the 2015 Criminal Procedure Code contains provisions that lay the foundation for control, such as Article 8 (guaranteeing human rights) and Article 99 (on the evidentiary value of electronic data).[43] These provisions provide a legal basis for controlling the evidentiary value of electronic data, including outputs from AI systems. Accordingly, all evidence must be collected, examined, and assessed in accordance with legal procedures, meaning that AI cannot automatically become a valid source of evidence without human verification. This approach reflects the spirit of the "human supervision" principle and ensures that the digitisation of justice maintains transparency, fairness, and humanity – the core values of the rule of law state.

39 Regulation (EU) 2024/1689 (n 3).

40 OHCHR, 'Human Rights Due Diligence for Digital Technology Use - Guidance of the Secretary-General: Practical Guide' (*United Nations Human Rights*, 30 September 2025) <https://www.ohchr.org/en/documents/tools-and-resources/human-rights-due-diligence-digital-technology-use-guidance> accessed 24 October 2025.

41 OECD, 'AI Principles' (2024) <https://www.oecd.org/en/topics/sub-issues/ai-principles.html> accessed 24 October 2025.

42 Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (5 September 2024) CETS 225 <https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence> accessed 24 October 2025.

43 Law of the Socialist Republic of Vietnam No 101/2015/QH13 (n 7).

4.4. The Principle of a Balanced Approach in Applying AI to Criminal Justice

The "balance" in the application of AI to criminal justice is not merely about finding a middle ground between technological efficiency and the protection of human rights, but rather a dynamic legal mechanism that ensures the values of the rule of law are maintained throughout the innovation process.[44] Balance is understood as a continuous legal balancing process, in which each stage of the proceedings (investigation, prosecution, trial, and enforcement) must be re-evaluated for the lawfulness, necessity, and proportionality of applying new technologies, over time and in specific contexts.[45]

The concept of dynamic balance originates from how the European Court of Human Rights (ECtHR) interprets Article 8 of the European Convention on Human Rights (ECHR) with the criteria of lawfulness, necessity, proportionality, and the "fair balance". In the case of *S. and Marper v. the United Kingdom*, the ECtHR considered the comprehensive and indefinite retention of biometric data from individuals not convicted of a crime to be unnecessary in a democratic society due to the lack of time limits and insufficient safeguards.[46] This implies that the level of intervention must be flexible in response to technological risks and be subject to periodic review. In *Big Brother Watch and Others v. the United Kingdom,* the ECtHR demanded "end-to-end safeguards" (including selection criteria, independent approval, and continuous monitoring) for mass data collection,[47] indicating that proportionality is not static but evolves alongside the risks and the intensity of actual monitoring.[48]

The EU's 2024 AI Act incorporates this concept into law through a lifecycle, risk-based governance model: "high-risk" AI systems in law enforcement and judicial operations must undergo periodical assessments, ensure transparency and explainability, and maintain "human supervision."[49] This means that humans have the authority to intervene, adjust, or suspend the system's operation when necessary. Simultaneously, the Council of Europe's AI Framework Convention requires countries to maintain a risk-based and rights-based balancing mechanism with safeguards throughout the entire lifecycle, thereby establishing a "dynamic balance" standard that encourages innovation without exceeding the boundaries of human rights and the rule of law.[50]

44 Jonida Milaj, 'Privacy, Surveillance, and the Proportionality Principle: The Need for a Method of Assessing Privacy Implications of Technologies Used for Surveillance' (2016) 30(3) International Review of Law, Computers & Technology 115. doi:10.1080/13600869.2015.1076993.

45 Mantelero (n 22; 23).

46 *S and Marper* (n 14).

47 *Big Brother Watch and Others* (n 27).

48 Kristina Trykhlib, 'The Principle of Proportionality in the Jurisprudence of the European Court of Human Rights' (2020) 4 EU and Comparative Law Issues and Challenges Series (ECLIC) 128. doi:10.25234/eclic/11899.

49 Regulation (EU) 2024/1689 (n 3).

50 Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (n 42).

In Vietnam, the principle of balance has started to emerge within the personal data protection framework. Decree 13/2023/ND-CP requires data processing to be purpose-driven, minimized, and subject to retention limits, while the 2025 Personal Data Protection Law establishes impact assessments for high-risk activities - requirements that can be directly integrated into each stage of legal proceedings to repeatedly verify the lawfulness, necessity, and proportionality of using AI.[51] Combined with Articles 8 and 99 of the 2015 Criminal Procedure Code,[52] a dynamic balancing mechanism can be designed on a case-by-case and time-bound basis to ensure that AI is a controlled support tool, not a replacement for human judicial judgment.

5 THE EU'S EXPERIENCE IN ENSURING THE BALANCE BETWEEN TECHNOLOGICAL INNOVATION AND PERSONAL DATA PROTECTION IN CRIMINAL JUSTICE

5.1. The EU's Legal Framework for Ensuring the Balance between Technological Innovation and Personal Data Protection

The experience of the EU shows that ensuring the balance between technological innovation and personal data protection is underpinned by a multi-layered legal framework, in which the principles of human rights and technological innovation are consistently integrated at the constitutional level. The European Union Charter of Fundamental Rights recognises the right to personal data protection as an independent right (Article 8),[53] which forms the foundation for the GDPR, LED, and AI Act– three legal pillars that directly regulate the relationship between technology and human rights in the field of criminal justice.

(i) The GDPR establishes core principles for data processing, including: lawfulness, transparency, purpose limitation, data minimisation, and accountability. Articles 5(2) and 35 of the GDPR require conducting a DPIA for high-risk activities, especially when applying new technologies in investigations or adjudication.[54]

(ii) The Directive (EU) 2016/680 (LED) extends the principles of GDPR to the criminal field, allowing flexibility in restricting certain rights (such as the right to access or object) when necessary for the proceedings. But it still requires the principles of lawfulness, necessity, proportionality, and independent oversight mechanisms.[55]

51 Decree of the Socialist Republic of Vietnam No 13/2023/ND-CP (n 8); Law of the Socialist Republic of Vietnam No 91/2025/QH15 (n 9).
52 Law of the Socialist Republic of Vietnam No 101/2015/QH13 (n 7).
53 Charter of Fundamental Rights of the European Union [2012] OJ C 326/391.
54 Regulation (EU) 2016/679 (n 5).
55 Directive (EU) 2016/680 (n 4).

(iii) The AI Act marks a new step in EU law by treating AI systems in the judicial, law enforcement, and national security fields as a "high-risk" group. The AI Act requires these systems to comply with transparency, training data governance, periodic risk assessments, and human oversight mechanisms, ensuring that judicial decisions are not "automated" but remain subject to human control.[56]

The three abovementioned documents do not stand separately but operate as mutually reinforcing normative layers, forming a dynamically balanced legal mechanism in which technology is developed within the framework of human rights and the rule of law. This system helps the EU to shift its approach from "protection against technology" to "responsible technological governance". The EU does not consider the right to personal data protection as a barrier but a necessary condition for legitimate and sustainable innovation. Based on that legal foundation, the EU has developed numerous enforcement and monitoring mechanisms to maintain a balance between technological innovation and human rights in judicial practice, especially at the member-state level.

5.2. Balancing Mechanisms through Impact Assessment and Risk Control

Based on the multi-layered legal framework established at the Union level, the EU has developed a series of preventive and adaptive enforcement mechanisms to maintain the balance between technological innovation and human rights protection throughout the entire lifecycle of data and AI systems. The focus of this mechanism lies in impact assessment and risk control—legal tools that help translate abstract human rights principles into concrete, measurable, and monitorable operational procedures in judicial practice.

A highlight of the EU experience is the prevention of legal risks and the protection of human rights from the technological design stage. It is based on the principle that innovation is legitimate only when carried out within predictable, controllable limits, ensuring that human rights are not compromised for the sake of efficiency.[57] According to Article 35 of the GDPR, data controllers must conduct a data protection impact assessment (DPIA) where processing activities are "likely to pose a high risk" to the rights and freedoms of individuals, in particular where "systematic surveillance on a large scale" or "new technologies are applied".[58] DPIA is not only a technical procedure, but a legal balancing mechanism that requires data processors to anticipate, model, and mitigate risks before deploying technology.

56 Regulation (EU) 2024/1689 (n 3).

57 EDPS, 'Guidelines on Assessing the Proportionality of Measures that Limit the Fundamental Rights to Privacy and to the Protection of Personal Data' (24 October 2025) <https://www.edps.europa.eu/data-protection/our-work/publications/guidelines/edps-guidelines-assessing-proportionality-measures> accessed 24 October 2025.

58 Regulation (EU) 2016/679 (n 5).

In the field of justice and security, Directive (EU) 2016/680 (LED) specifies this mechanism in Articles 27 – 31.[59] It requires competent authorities to clearly define the legal basis for the collection, processing, and storage of data, to ensure that the processing purposes are clearly defined and that the storage period does not exceed the needs of the investigation, prosecution, or adjudication. LED also demands the establishment of an independent monitoring system led by the national data protection authority. This is an ex post control mechanism that helps maintain the balance between judicial power and the legal framework, ensuring the rule of law.

The combination of ex ante and ex post control creates a continuous balancing cycle in the governance of judicial technology. This approach is reinforced by the Court of Justice of the European Union (CJEU) through case law. Typically, in *Digital Rights Ireland Ltd. v. Minister for Communications*, CJEU held that the indiscriminate retention of telecommunications data from all citizens violates the principle of proportionality, as there are no clear limits on the scope, duration, and mechanism of access to the data.[60] Subsequently, *Tele2 Sverige AB v. Watson* affirmed that any data monitoring mechanism must be based on "objective criteria" and "independent prior review", thereby establishing a new legal standard for controlling technological risks in the justice sector.[61] Case law is fundamental to the entire EU legal framework on AI and data in the justice sector.

Thus, the EU model does not concentrate on "preventing" technology, but on creating a legal framework for responsible innovation. DPAI, FRIA, and the independent monitoring mechanism are not only technical tools but also institutional manifestations of the principles of balancing judicial efficiency and the protection of human rights—an important experience for Vietnam in perfecting the legal framework for applying AI in criminal proceedings.

5.3. Enforcement Mechanism and the Role of Independent Monitoring Bodies

If the principles of lawfulness, necessity, and proportionality establish "theoretical standards" for data processing, then enforcement mechanisms and independent monitoring bodies serve to ensure these standards are implemented effectively, transparently, and verifiably. In the EU system, this enforcement structure is designed based on three complementary pillars: (i) an independent monitoring body with substantive enforcement powers; (ii) a technical–legal oversight mechanism along the life cycle of high-risk

59 Directive (EU) 2016/680 (n 4).

60 Joined Cases C-293/12 and C-594/12 *Digital Rights Ireland Ltd v Minister for Communications, Marine and Natural Resources and Others and Kärntner Landesregierung and Others* [2014] ECJ ECLI:EU:C:2014:238 <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:62012CJ0293> accessed 24 October 2025.

61 Joined Cases C-203/15 and C-698/15 *Tele2 Sverige AB v Post- och telestyrelsen and Secretary of State for the Home Department v Tom Watson and Others* [2016] ECJ ECLI:EU:C:2016:970 <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:62015CJ0203> accessed 24 October 2025.

technologies; and (iii) a unified coordination mechanism across the Union to maintain common rule of law standards.

(1) *Independent monitoring body – the heart of the balancing mechanism.* This principle was established in the case of Commission v. Germany, in which the CJEU held that any government intervention in the activities of a data protection authority violates the duty of absolute independence.[62] This was reinforced in the Schrems II case, where the Court stressed that the monitoring authority must be able to proactively suspend data processing when there is a risk of a fundamental right being violated.[63] This provision is codified in Article 52 GDPR and Article 41 LED, requiring Data Protection Authorities (DPAs) to have organisational autonomy, budget, and enforcement powers.

(2) *Lifecycle technical oversight mechanism – ensuring balance in the innovation process.* To ensure consistency and implementation, the European Commission has established the AI Office. This agency has the authority to request that AI developers provide technical dossiers, testing information, incident reports, and evidence of human oversight to ensure compliance with the obligations under the AI Act.[64] This mechanism shifts the concept of "balance" from a static state to a dynamic one, in which all AI applications in the justice sector are not only approved once but also periodically reviewed as technology and risk levels evolve. This is a manifestation of evolutionary balance, in which the legality and the protection of human rights of the technology are continuously re-examined throughout its operational lifecycle.

(3) *Union-wide consistency and coordination mechanism – ensuring systemic balance.* In addition to national oversight, the EU has established a consistency mechanism to maintain uniformity in data protection. Under Articles 63 – 65 of the GDPR, if DPAs have different views on a cross-border case, the European Data Protection Board (EDPB) can issue a binding decision that the parties must comply with.[65] This mechanism creates horizontal balance among Member States, prevents legal fragmentation, and ensures that all individuals in the EU enjoy equal protection from technology-related risks, irrespective of their residence or the location of data processing.

62 *Uzun v Germany* (n 25).

63 Case C-311/18 *Data Protection Commissioner v Facebook Ireland and Maximillian Schrems II* [2020] ECJ ECLI:EU:C:2020:559 <https://eur-lex.europa.eu/legal-content/NL/TXT/?uri=ecli:ECLI:EU:C:2020:559> accessed 24 October 2025.

64 'European AI Office' (*European Commission: Shaping Europe's Digital Future*, 2025) <https://digital-strategy.ec.europa.eu/en/policies/ai-office> accessed 24 October 2025.

65 EDPB, 'Guidelines 03/2021 on the Application of Article 65(1)(a) GDPR' (adopted 13 April 2021) <https://www.edpb.europa.eu/our-work-tools/documents/public-consultations/2021/guidelines-032021-application-article-651a-gdpr_en> accessed 24 October 2025.

In the field of criminal justice, the three-layered mechanism has profound legal implications. The combination of independent oversight and technical monitoring ensures that procedural efficiency cannot be achieved at the expense of human rights. This approach allows the EU to implement the principle of "justice must not be automated": technology can assist in judgment, but only humans can make the final judicial decisions.

The above comparison shows that the main differences lie in institutional standardisation and risk-governance mechanisms. The EU has developed a three-tier legal framework (GDPR, LED, AI Act) with independent oversight, while Vietnam is still forming its basic structure, lacking both an independent data authority and mandatory impact assessments. The EU's experience offers guidance for Vietnam to balance innovation with human rights protection in criminal justice. Although the EU is regarded as a pioneering jurisdiction in data protection and technological risk governance, its model also has limitations, such as the implementation of the 2024 AI Act, which has generated debate over its enforceability, particularly given the significant differences in infrastructure readiness and supervisory capacity among Member States. In addition, the Schrems II judgment[66] showed that the EU continues to face difficulties in ensuring cross-border data security and maintaining consistent standards with international partners. These limitations indicate that the EU's experience does not constitute a "perfect model," but rather serves as a reference point for Vietnam as it develops a balanced mechanism suited to its national context.

6 LEGAL STATUS AND CHALLENGES IN VIETNAM IN ENSURING A BALANCE BETWEEN TECHNOLOGICAL INNOVATION AND PERSONAL DATA PROTECTION IN CRIMINAL JUSTICE

The process of digital transformation in the justice sector in Vietnam creates many opportunities for innovation, but also poses an urgent need to ensure a balance between technological efficiency and the right to protect personal data. Although the fundamental legal framework, such as the 2015 Criminal Procedure Code, Decree 13/2023/ND-CP, and the 2025 Law on Personal Data Protection, has been issued, the current legal system still lacks a mechanism to incorporate, monitor, and control technological risks in criminal justice.

6.1. Current Status of Vietnamese Laws

The 2015 Criminal Procedure Code, for the first time, recognises electronic data as a source of evidence (Articles 87 and 99) and allows the use of digital means in the process of collecting, examining, and evaluating evidence.[67] This provision marked a turning point, shifting from traditional to digital thinking in litigation and paving the way for the

66 Case C-311/18 (n 63).

67 Law of the Socialist Republic of Vietnam No 101/2015/QH13 (n 7).

integration of automated analysis tools and decision-support systems in investigation, prosecution, and adjudication. Additionally, Decree No. 13/2023/ND-CP on Personal Data Protection and the 2025 Law on Personal Data Protection (effective from 1 January 2026) have formed a unified legal basis for the right to protect personal data.[68] This Decree stipulates seven foundational principles for processing data, including lawfulness, transparency, purpose limitation, and data minimisation. In particular, Articles 24 and 25 impose an obligation to conduct an impact assessment for high-risk data processing activities, laying the groundwork for a risk-prevention mechanism similar to the Data Protection Impact Assessment (DPIA) in EU Law.

The 2025 Law on Personal Data Protection expands the scope of protection, establishing the right of individuals to know, object, and request the termination of data processing, and clearly stipulates accountability for organisations, state agencies, and enterprises when processing data in judicial and administrative activities. However, the current Vietnamese legal framework still reveals some limitations as follows:

First, the regulations governing data processing in criminal justice activities remain fragmented, and a coordination mechanism among the Investigation Agency, the Procuracy, the Court, and the Data Management Agency has not yet been established. Meanwhile, the EU model clearly stipulates the role of independent Data Protection Authorities in supervising the processing of data for the purposes of investigation and prosecution.

Second, there are no specific regulations on accountability and risk assessment when applying AI in the justice sector. Current documents are mainly at the framework level and do not identify the agency responsible for appraising, approving, or supervising AI systems used in legal proceedings. This poses a potential risk of "automating judicial power," when the results generated by AI systems (for example: risk scoring, behavioural analysis) can directly influence investigation or adjudication decisions without being fully verified by humans.[69]

Third, the lack of an independent monitoring body for personal data protection–a "focal pillar" of the EU model. Currently, data management in Vietnam is dispersed between the Ministry of Public Security and the Ministry of Science and Technology, leading to overlapping responsibilities and a lack of a unified judicial and administrative monitoring mechanism.

Finally, the institutional and digital capacities of the judiciary team remain limited compared to practical requirements. The understanding and application of data protection principles, such as "necessary", "proportionate", and "human supervision",

68 Decree of the Socialist Republic of Vietnam No 13/2023/ND-CP (n 8); Law of the Socialist Republic of Vietnam No 91/2025/QH15 (n 9).

69 Mantelero (n 22; 23).

remain at the level of awareness and have not yet become mandatory standards in investigation and adjudication.

Therefore, although the initial legal foundation for ensuring the right to personal data protection has been established, the application of AI to criminal justice in Vietnam still requires integrated improvement, grounded in human rights and the control of high-risk technology. This involves a combination of institutional reform, the establishment of independent monitoring bodies, and capacity-building for legal proceedings and conducting bodies to achieve a dynamic balance between technological innovation and human rights protection. Recent developments provide important indications of rising risks and the urgent need for data protection mechanisms in the context of judicial digitalisation. In Ho Chi Minh City, a system of 31 AI-integrated cameras detected more than 3,100 violations within just over one month of operation, illustrating that behavioural data are being collected on a large and continuous scale.[70] On the other hand, the Ministry of Public Security dismantled a scheme involving the sale of nearly 56 million personal data records,[71] reflecting the severe vulnerability of individuals to data breaches. At the same time, the Ministry of Justice rose six places in the 2023 ministerial-level Digital Transformation Index.[72] It showed efforts to enhance the digital capacity of the justice sector—but also signalling the urgent need for corresponding risk-control mechanisms as technological applications continue to expand.

6.2. Challenges of Vietnam in Ensuring the Balance between Technological Innovation and the Right to Personal Data Protection in Criminal Justice

First, Vietnam currently lacks a specialised legal framework regulating the use of AI in criminal justice. The 2015 Criminal Procedure Code only recognises electronic data as a source of evidence (Articles 87 and 99). It does not regulate legal conditions, evidentiary value, or exclusion rules for data generated by AI systems. Litigation practice shows that there are challenges in authenticating, collecting evidence, and assessing the value of

70 Vu Phuong, 'AI-Enabled Cameras Scanning the Streets of Ho Chi Minh City Detect More than 3,100 Traffic Violations' *Báo Thanh Niên* (Ho Chi Minh city, 5 October 2025) [in Vietnamese] <https://thanhnien.vn/camera-ai-quet-duong-pho-tphcm-phat-hien-hon-3100-truong-hop-vi-pham-giao-thong-185251005140705779.htm> accessed 21 November 2025.

71 Hai Lan, 'A Trafficking Ring Trading Nearly 56 Million Personal Data Records Was Dismantled' *Báo Công An Nhân Dân* (Hanoi City, 21 February 2025) [in Vietnamese] <https://cand.com.vn/Ho-so-Interpol/triet-pha-duong-day-mua-ban-gan-56-trieu-du-lieu-ca-nhan-i759592/> accessed 21 November 2025.

72 An Nhu, 'The Ministry of Justice Rose Six Places in the 2023 Digital Transformation Index' (*Bộ Tư pháp*, 6 February 2025) [in Vietnamese] <https://moj.gov.vn/qt/tintuc/Pages/hoat-dong-cua-lanh-dao-bo.aspx?ItemID=6810> accessed 21 November 2025.

electronic evidence, especially when AI technology is applied in activities of identifying, predicting, or analysing behaviour.[73]

Second, the DPIA mechanism is only stipulated at the principle-based level, without specific technical guidance on risk criteria, appraisal procedures, or approval agencies. Meanwhile, EU standards (Article 35 GDPR; EDPB Guidelines 4/2020) require that all high-risk data processing activities, especially in the justice and security fields, must undergo independent audits. The absence of this mechanism hinders Vietnam's ability to ensure a "dynamic balance" between technological innovation and human rights.

Third, the institutional capacity and human resources of the judiciary for auditing, explaining, and monitoring the operation of AI systems remain limited. Investigators, prosecutors, and judges are not fully equipped with the skills to understand, criticise, or verify the objectivity of algorithms. Meanwhile, studies by OHCHR[74] and the EU Agency for Fundamental Rights[75] have warned that the application of AI in law enforcement can lead to bias, discrimination, and privacy violations without human oversight and transparent accountability mechanisms.

Fourth, data sharing and interconnection between judicial agencies are limited and lack protection standards. Data is stored in a decentralised manner by each agency. There are no technical standards for anonymisation, encryption, or limitation of use. This not only affects the effectiveness of coordination in legal proceedings but also increases the risk of leakage or misuse.

Fifth, the standard for the right to personal data protection is not clearly defined in the Constitution. The 2013 Constitution (Article 21) only recognises the right to privacy and personal confidentiality, but does not recognise the right to personal data protection as an independent right. Meanwhile, international standards consider this right a "digital personality right" that helps regulate structural risks in the AI era. Without being constitutionalised, specialised laws lack a constitutional basis to establish an effective balancing mechanism between innovation and human rights protection.[76]

Overall, the above challenges show that Vietnam is only in the early stages of developing a "dynamic balance" model between technological innovation and the protection of personal

73 Vo Minh Tuan, 'Difficulties and Barriers Regarding Electronic Data in the 2015 Criminal Procedure Code' [7 February 2021] Tạp chí Tòa án nhân dân điện tử [in Vietnamese] <https://tapchitoaan.vn/kho-khan-vuong-mac-ve-du-lieu-dien-tu-trong-bo-luat-to-tung-hinh-su-nam-2015> accessed 24 October 2025.

74 OHCHR, 'The Right to Privacy in the Digital Age: Report of the United Nations High Commissioner for Human Rights' (A/HRC/48/31, 13 September 2021) <https://www.ohchr.org/en/documents/thematic-reports/ahrc4831-right-privacy-digital-age-report-united-nations-high> accessed 24 October 2025.

75 FRA, *Facial Recognition Technology: Fundamental Rights Considerations in EU Law Enforcement* (Publications Office of the EU 2023).

76 OHCHR (n 74); Council of Europe, *Convention 108+* (n 18).

data rights. To move towards a digital justice model that ensures human rights, Vietnam must shift from a reactive legal framework to a proactive one, in which all AI applications in criminal justice are grounded in lawfulness, necessity, proportionality, accountability, and human oversight. Moreover, reports from international organisations also indicate structural technological risks associated with the use of AI in judicial activities. The OHCHR report warns of algorithmic bias in facial recognition and crime-prediction systems, particularly in relation to women and minority groups.[77] Similarly, FRA reports high misidentification rates and the risk of biometric data leakage when law enforcement authorities store data on a large scale without adequate access controls.[78] These risks underscore the urgent need to design appropriate oversight and impact-assessment mechanisms as Vietnam expands the use of AI in criminal justice.

Compared with the EU, Vietnam shares the same core values of personal data protection but differs in institutionalisation and enforcement. The EU mandates Data Protection and Fundamental Rights Impact Assessments (DPIA/FRIA) and empowers independent supervisory authorities, while Vietnam only outlines general principles and distributes authority across multiple agencies. This gap reflects not only differences in capacity but also the need for selective and context-appropriate legal adaptation.

7 RECOMMENDATIONS TO IMPROVE VIETNAM'S LEGAL SYSTEM FROM THE EUROPEAN UNION'S EXPERIENCE

Based on the EU's experience, this section proposes directions for improving the law and policy implications to help Vietnam ensure a "dynamic balance" between technological efficiency and human rights. However, it is necessary to understand that strengthening personal data protection in criminal justice is a long-term undertaking that requires continuous and incremental institutional adjustment. As a developing country, Vietnam must simultaneously undertake many other important strategic tasks, particularly economic development; therefore, legal reforms in this area need to be implemented through a multi-stage roadmap, prioritising the improvement of the legal framework and enforcement capacity, and only subsequently moving toward independent oversight mechanisms and adequate technological infrastructure.

(1) *Leverage the right to personal data protection to constitutional standards.* Vietnam's 2013 Constitution addresses this right only in Article 21 on privacy and correspondence. Elevating it to a constitutional or statutory level would (i) reinforce the rule of law in the digital era by subjecting technological activities in justice to constitutional oversight, and (ii) provide a legal basis for balancing rights with public interests and crime prevention. As a first step, this could be detailed in the 2025 Law

77 OHCHR (n 74).
78 FRA (n 75).

on Personal Data Protection's guiding documents, paving the way for future constitutional reform.

(2) *Complete the Criminal Procedure Code on "digital evidence detected or found with the support of AI".* Currently, the 2015 Criminal Procedure Code only stipulates electronic data as a type of evidence (Article 99) without any legal standards for evidence detected or found with the support of AI. Therefore, it is essential to add a specialised sub-section on AI evidence to the Criminal Procedure Code, including: conditions of acceptance (model transparency, explainability, and independent verification); exclusion rules for evidence violating the right to personal data protection; the obligation to keep technical traces (traceability); and the right to challenge the algorithm of the defender.

(3) *Add a data impact assessment mechanism (DPIA/FRIA).* Vietnam could adopt an EU-style mechanism requiring mandatory DPIAs for all AI projects in investigation, adjudication, and enforcement, with both ex-ante and ex-post reviews by a data oversight body. In the short term, this could be specified in the Decree guiding the 2025 Law on Personal Data Protection, incorporating a verification process akin to "professional approval" in healthcare. Vietnam should also establish an independent National Data Protection Authority, empowered to supervise data handling in criminal proceedings in coordination with the Supreme People's Procuracy—similar to the EU's EDPB—to enhance transparency and public trust in judicial digitalisation.

(4) *Establish a separate legal framework for AI in the criminal justice system.* Vietnam could develop a principles-based legal framework for judicial AI covering three areas: (i) defining judicial AI, (ii) classifying risks and corresponding liabilities, and (iii) creating a regulatory sandbox for technologies in adjudication and enforcement. This could be piloted at the Supreme People's Court and the Ministry of Public Security, following the State Bank's fintech sandbox model. At the same time, a national human rights–oriented AI policy for criminal justice that defines the scope of AI use across investigation, prosecution, adjudication, and enforcement, and establishes ethical, technical, and legal safeguards. Vietnam may adopt the EU's "Human Rights Impact Assessment" model to evaluate and disclose data-sensitive technologies. This is feasible given the foundations established under Decree No. 13/2023/NĐ-CP and the 2025 Law on Personal Data Protection, which already require DPIAs.

(5) *Refine the policy of training, fostering digital capacity, and technology ethics for judicial staff.* Vietnam should establish a continuous training program on "judicial data governance" and, through international cooperation within the EU-ASEAN or UNESCO frameworks, exchange experiences and methods for technology verification. The investment in human capacity will determine the level of success of the process of "humanising digital justice".

(6) *The State can encourage technology enterprises to participate in research, testing, and the development of judicial support tools,* such as electronic record management systems, digital evidence analysis, or virtual assistants for conducting legal proceedings. But these tools must be subject to independent verification of data risks and human rights impacts. This approach encourages innovation and ensures that the private sector cannot dominate or undermine the independence of the judicial activities.

(7) *Promote international cooperation on judicial technology.* Vietnam should participate in international initiatives, such as the Global Partnership on AI (GPAI) or the OECD.AI Policy Observatory program. This can support Vietnam in approaching global standards on responsible AI and personal data protection. Additionally, it is necessary to promote bilateral judicial dialogue with the EU within the framework of implementing the EVFTA and EVIPA, to learn about mechanisms to ensure data rights in the cross-border judicial environment.

The above analysis shows that the EU has achieved balance through three institutional pillars: the constitutional right to data protection, mandatory impact assessments, and independent lifecycle supervision. Vietnam is gradually approaching this model through the 2025 Law on Personal Data Protection and the revision of the Criminal Procedure Code. Incorporating these elements domestically will promote "responsible innovation," a crucial step toward safeguarding human rights in the digital justice system.

8 CONCLUSIONS

The rapid development of AI is reshaping criminal justice, especially in evidence collection, processing, and evaluation. Along with the obvious benefits in efficiency, speed, and analytical capabilities, this technology also poses challenges for human rights protection, with the right to data protection becoming a focus of contemporary legal debate.

The experience of the EU shows that "balancing" between technological innovation and human rights protection is not a choice between two opposite poles, but a continuous process of legal governance. The EU has built a systemic regulatory model – combining a constitutional basis for the right to personal data, risk assessment and accountability mechanisms, risk–based management, and independent monitoring and judicial control. For Vietnam, the promulgation of the 2025 Law on Personal Data Protection and the adoption of digital transformation policies in criminal justice mark an essential step forward. However, to achieve a sustainable balance between technological innovation and the protection of human rights, it is necessary to establish a comprehensive legal framework that clearly sets out the scope, limits, and control mechanisms for the use of AI in criminal proceedings.

REFERENCES

1. An N, 'The Ministry of Justice Rose Six Places in the 2023 Digital Transformation Index' (*Bộ Tư pháp*, 6 February 2025) [in Vietnamese] <https://moj.gov.vn/qt/tintuc/Pages/hoat-dong-cua-lanh-dao-bo.aspx?ItemID=681> accessed 21 November 2025
2. Daly A and others, 'Artificial Intelligence, Governance and Ethics: Global Perspectives' (Research Paper no 2019-15, The Chinese University of Hong Kong Faculty of Law 2019) doi:10.2139/ssrn.3414805
3. Hai L, 'A Trafficking Ring Trading Nearly 56 Million Personal Data Records Was Dismantled' *Báo Công An Nhân Dân* (Hanoi City, 21 February 2025) [in Vietnamese] <https://cand.com.vn/Ho-so-Interpol/triet-pha-duong-day-mua-ban-gan-56-trieu-du-lieu-ca-nhan-i759592/> accessed 21 November 2025
4. Mantelero A, 'AI and Big Data: A Blueprint for a Human Rights, Social and Ethical Impact Assessment' (2018) 34(4) Computer Law & Security Review 754. doi:10.1016/j.clsr.2018.05.017
5. Mantelero A, 'The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, Legal Obligations and Key Elements for a Model Template' (2024) 54 Computer Law & Security Review 106020. doi:10.1016/j.clsr.2024.106020
6. Milaj J, 'Privacy, Surveillance, and the Proportionality Principle: The Need for a Method of Assessing Privacy Implications of Technologies Used for Surveillance' (2016) 30(3) International Review of Law, Computers & Technology 115. doi:10.1080/13600869.2015.1076993
7. Stahl BC and others, 'Artificial Intelligence for Human Flourishing – Beyond Principles for Machine Learning' (2021) 124 Journal of Business Research 374. doi:10.1016/j.jbusres.2020.11.030
8. Trykhlib K, 'The Principle of Proportionality in the Jurisprudence of the European Court of Human Rights' (2020) 4 EU and Comparative Law Issues and Challenges Series (ECLIC) 128. doi:10.25234/eclic/11899
9. Vo MT, 'Difficulties and Barriers Regarding Electronic Data in the 2015 Criminal Procedure Code' [7 February 2021] Tạp chí Tòa án nhân dân điện tử [in Vietnamese] <https://tapchitoaan.vn/kho-khan-vuong-mac-ve-du-lieu-dien-tu-trong-bo-luat-to-tung-hinh-su-nam-2015> accessed 24 October 2025
10. Vu P, 'AI-Enabled Cameras Scanning the Streets of Ho Chi Minh City Detect More than 3,100 Traffic Violations' *Báo Thanh Niên* (Ho Chi Minh city, 5 October 2025) [in Vietnamese] <https://thanhnien.vn/camera-ai-quet-duong-pho-tphcm-phat-hien-hon-3100-truong-hop-vi-pham-giao-thong-18525100514070577 9.htm> accessed 21 November 2025
11. Wachter S and Mittelstadt B, 'A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI' (2019) 2 Columbia Business Law Review 494

AUTHORS INFORMATION

Phuong Anh Nguyen
Master of Law, PhD candidate, Faculty of Criminal Law, Hanoi Law University, Hanoi, Vietnam
phuonganhlawb@gmail.com
https://orcid.org/0009-0002-3555-3003
Corresponding author, solely responsible for the manuscript preparing.

Competing interests: No competing interests were disclosed.

Disclaimer: The author declares that their opinion and views expressed in this manuscript are free of any impact of any organizations.

RIGHTS AND PERMISSIONS

EDITORS

ABOUT THIS ARTICLE

Cite this article
Nguyen PA, 'Artificial Intelligence in Criminal Justice: Balancing Technological Innovation and Personal Data Protection Rights. A Comparative Legal Study between the European Union and Vietnam' (2025) 8(Spec) Access to Justice in Eastern Europe 233-58 <https://doi.org/10.33327/AJEE-18-8.S-r000161>

DOI: https://doi.org/10.33327/AJEE-18-8.S-r000161

Keywords: *artificial intelligence, criminal justice, right to personal data, human rights, European Union, Vietnam.*

DETAILS FOR PUBLICATION

Date of submission: 26 Oct 2025
Date of acceptance: 08 Dec 2025
Online First publication: 19 Dec 2025
Last Published: 30 Dec 2025
Whether the manuscript was fast tracked? - No
Number of reviewer report submitted in first round: 3 reports
Number of revision rounds: 1 round with minor revisions

Technical tools were used in the editorial process:
Plagiarism checks - Turnitin from iThenticate https://www.turnitin.com/products/ithenticate/
Scholastica for Peer Review https://scholasticahq.com/law-reviews

AI DISCLOSURE STATEMENT

The author confirms that AI technologies have only been used to enhance language clarity and grammar. No AI tools were used to generate ideas, structure arguments, analyze data, or produce conclusions.

АНОТАЦІЯ УКРАЇНСЬКОЮ МОВОЮ

Дослідницька стаття

ШТУЧНИЙ ІНТЕЛЕКТ У КРИМІНАЛЬНОМУ СУДОЧИНСТВІ: БАЛАНС МІЖ ТЕХНОЛОГІЧНИМИ ІННОВАЦІЯМИ ТА ПРАВОМ НА ЗАХИСТ ПЕРСОНАЛЬНИХ ДАНИХ. ПОРІВНЯЛЬНО-ПРАВОВЕ ДОСЛІДЖЕННЯ ПІДХОДІВ ЄВРОПЕЙСЬКОГО СОЮЗУ ТА В'ЄТНАМУ

Фуонг Ан Нгуєн

АНОТАЦІЯ

Вступ. *Швидкий розвиток штучного інтелекту сильно змінив різні аспекти суспільного життя, зокрема систему кримінального судочинства. У кримінальному провадженні збір та обробка біометричних, поведінкових та емоційних даних може загрожувати праву на приватність, презумпції невинуватості та праву на справедливий суд. У цьому дослідженні розглядається взаємозв'язок між технологічними інноваціями та захистом персональних даних у кримінальному судочинстві Європейського Союзу та В'єтнаму. За допомогою порівняльно-правового аналізу Загального регламенту про захист даних (GDPR) ЄС, Директиви про правоохоронну діяльність та Акту ЄС про ШІ 2024 року разом із правовою базою В'єтнаму, у статті визначено ключові сфери конвергенції, розбіжностей та регуляторних прогалин.*

Методи. *У дослідженні використовується порівняльно-правовий аналіз у поєднанні з підходом, що ґрунтується на правах людини, для уточнення взаємозв'язку між технологічними інноваціями та правом на захист персональних даних у кримінальному судочинстві. Джерелами інформації є правові бази Європейського Союзу та В'єтнаму, судова практика та звіти органів ЄС та Організації Об'єднаних Націй.*

Результати та висновки. *Метою дослідження є встановлення фундаментальних правових принципів, які забезпечують баланс між технологічними інноваціями та захистом права на персональні дані у кримінальному судочинстві — підхід, якому приділяється недостатня увага у В'єтнамі. На основі цього було запропоновано правові реформи, спрямовані на створення системи «цифрового правосуддя, орієнтованого на права людини», яка гарантуватиме, що цифровізація та застосування штучного інтелекту в системі правосуддя не лише підвищать операційну ефективність, але й зможуть зміцнити верховенство права та захистити людей.*

Ключові слова: *штучний інтелект, кримінальне судочинство, право на захист персональних даних, права людини, Європейський Союз, В'єтнам.*

Access to Justice in Eastern Europe
ISSN 2663-0575 (Print)
ISSN 2663-0583 (Online)
http://ajee-journal.com

REVIEW ARTICLE

Review Article

LEGAL CHALLENGES RELATED TO CONTRACTUAL NEGOTIATIONS VIA AI TECHNOLOGIES: COMPARATIVE ANALYTICAL STUDY

Bashar Talal Momani*, Nasr Farid Hassan, Hosni Mahmoud AbdelDaiem AbdelSamad and Mohamed Elsayed Eldessouky

DOI:
https://doi.org/10.33327/AJEE-18-8.S-r000151

Date of submission: 28 Apr 2025
Date of acceptance: 02 Sep 2025
Online First publication: 27 Oct 2025
Last Published: 30 Dec 2025

ABSTRACT

Background: *Contractual negotiations conducted by artificial intelligence (AI) systems raise profound legal challenges, most notably the question of allocating civil liability for damages caused by their errors. This study, employing a comparative analytical methodology, reveals a significant regulatory gap in Arab jurisdictions—particularly Egypt and the United Arab Emirates—where legislation lacks explicit provisions governing such liability. In contrast, recent European Union initiatives, including the risk-based approach of the AI Act and the emerging framework of the AI Liability Directive, place primary emphasis on the accountability of developers and operators.*

Against this backdrop, the paper advocates for the development of a specialised Arab legal framework that draws inspiration from comparative models while preserving local specificities. Such a framework should include: a precise legal definition of

intelligent systems, concrete evidentiary mechanisms for fault attribution and liability distribution, the establishment of a dedicated supervisory authority, and the strengthening of insurance mechanisms as complementary safeguards.

Methods: *This study employs a comparative analytical method to examine civil liability for AI errors in contractual negotiations, focusing on tort and contractual theories under Egyptian and Emirati law, and contrasting them with recent EU developments—particularly the AI Liability Directive and the Data Act, which provide clearer guidance than the AI Act.*

Results and conclusions: *The comparative analysis yields three main results. First, there is a clear regulatory gap in Arab jurisdictions, which continue to rely on general civil law provisions without specialised rules for AI. Second, doctrinal differences between strict liability in the EU and the broader remedial approach in Arab systems complicate any direct transposition of European models. Third, evidentiary challenges remain central in both systems, as establishing fault and causation in AI-related harm is inherently complex.*

1 INTRODUCTION

Artificial intelligence (AI) is a real and effective technology whose applications, in some areas, exceed human capabilities. It is capable of reasoning, perception, problem-solving, and even autonomous learning. AI systems can integrate and utilise various advanced tools and devices to perform complex tasks efficiently.

However, the development of self-learning AI systems remains a major challenge. Many questions arise regarding their training, ethical use, and responsibility. At present, few mechanisms exist to adapt AI to specific cultural or linguistic contexts, particularly in environments where AI systems must recognise and respond to unique local characteristics.

To ensure the safe and ethical use of AI, it is essential to develop robust systems that promote transparency, accountability, and fairness. Such systems would facilitate automatic control while minimising potential risks.

This issue is particularly important in light of recent technological advancements, as AI increasingly influences contractual frameworks and civil liabilities, especially in regions where Arabic is the dominant language, such as Egypt and the United Arab Emirates. There is an urgent need for individuals to understand the basic concepts of AI and related technologies, especially in this rapidly evolving digital environment.

A review of Egyptian and Emirati legislation reveals significant shortcomings in addressing the legal challenges posed by AI-based negotiation mechanisms. Neither legal system includes explicit provisions recognising the legal status of AI systems or robots, nor do they provide clear definitions of the rights and obligations applicable to such artificial entities. Moreover, both jurisdictions lack a dedicated regulatory or supervisory body to oversee the operation of AI agents or to ensure accountability for civil liability arising from their role

in contractual negotiations. This stands in contrast to the more advanced European approach, which has moved toward establishing independent oversight bodies and imposing greater transparency obligations. In this regard, it is worth noting that the European Law Institute has issued a set of guiding principles on automated decision-making within the European Union.[1]

Within the European legal framework, it is essential to distinguish among three principal instruments. The EU Artificial Intelligence Act (AI Act), enacted as Regulation (EU) 2024/1689,[2] establishes a comprehensive risk-based legal regime for AI systems. It classifies such systems into four levels of risk—unacceptable, high, limited, and minimal—and imposes corresponding obligations on providers and deployers, including conformity assessments and transparency requirements.

The EU Data Act (Regulation (EU) 2023/2854)[3] focuses on enabling fair access to and use of data within the European data economy.[4] It forms a critical part of the legal foundation for smart contracts, especially in data-sharing agreements. These smart contracts are subject to essential legal requirements, such as auditability, access control, and conformity certification.

The AI Liability Directive[5] (AILD)—originally proposed to harmonise non-contractual civil liability rules for AI across the Union—seeks to simplify the claimant's burden of proof by facilitating access to evidence and introducing rebuttable presumptions of causality.[6] However, its legislative future remains uncertain.

1 Teresa Rodriguez de las Heras Ballell, *Guiding Principles for Automated Decision-Making in the EU* (ELI Innovation Paper, European Law Institute 2022).

2 Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules On Artificial Intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) [2024] OJ L 1689/1 <http://data.europa.eu/eli/reg/2024/1689/oj> accessed 20 April 2025. Detailed risk-based AI classification and obligations framework.

3 Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on Harmonised Rules on Fair Access to and Use of Data and amending Regulation (EU) 2017/2394 and Directive (EU) 2020/1828 (Data Act) [2023] OJ L 2854/1 <http://data.europa.eu/eli/reg/2023/2854/oj> accessed 20 April 2025. Aims to harmonize fair access to and use of data, including smart contract rules.

4 Sandra Wachter, Brent Mittelstadt and Luciano Floridi, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) 7(2) International Data Privacy Law 76. doi:10.1093/idpl/ipx005.

5 Proposal for a Directive of the European Parliament and of the Council on Adapting Non-Contractual Civil Liability Rules to Artificial Intelligence (AI Liability Directive) COM/2022/496 final (28 September 2022) <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52022PC0496> accessed 20 April 2025. Explores the EU's evolving framework for AI-related harm, including proposals for the AILD.

6 Andrea Bertolin, *Artificial Intelligence and Civil Liability: A European Perspective* (Study, European Parliament's Policy Department for Justice 2025) <https://www.europarl.europa.eu/thinktank/en/document/IUST_STU(2025)776426> accessed 25 July 2025.

The main research question guiding the study is: How can Arab regulatory frameworks bridge the gap in civil liability arising from AI-driven negotiations by leveraging the European model, while accounting for the region's legal specificities?

This study seeks to address the pressing legal challenges arising from the use of AI in contractual negotiations by formulating solutions consistent with prevailing legal traditions and legislative frameworks. Its objectives are fourfold: first, to emphasise the significance of the pre-contractual stage and underscore the legislator's duty to regulate it, particularly in determining the scope of civil liability associated with this phase; second, to examine the legal implications of AI in negotiation and the challenges it generates; third, to propose a balanced legal framework suitable for the responsible deployment of AI in contractual bargaining; and fourth, to explore the attribution of liability for harms caused by AI.

2 METHODOLOGY OF THE STUDY

This study adopts a comparative analytical approach, aiming to examine the legal framework governing civil liability for AI-related errors during the negotiation phase. The analysis focuses on the fundamental legal theories of tort and contractual liability as established in Arab civil laws, particularly those of Egypt and the UAE. This comparative approach is also employed to examine this situation in relation to recent legislative and judicial developments in the European Union, specifically the AI Liability Directive and the Data Act, which address these issues more directly than the AI Act.

3 AI CONCEPT AND RULE IN CONTRACTUAL NEGOTIATION

AI is a branch of computer science focused on developing systems that simulate human behaviour and decision-making with varying degrees of autonomy.[7] It operates through software or integrated devices such as robots and self-driving cars. AI processes structured, semi-structured, and unstructured data to analyse environments and solve complex problems. Its applications include digital assistants, facial recognition, and autonomous machines like drones. Due to their efficiency and adaptability, these technologies are increasingly utilised in medicine, economics, and defence.[8]

Contractual negotiation involves the exchange of proposals, opinions, studies, and legal consultations between parties aiming to reach a mutually beneficial agreement. It plays a crucial role in determining the terms and conditions of contracts, ensuring that the interests of all parties are balanced and that business transactions are successful. Negotiations

7 Arnaud Sée, 'La Régulation des Algorithmes: Un Nouveau Modèle de Globalisation?' (2019) 5 Revue Française de Droit Administratif 830.

8 Zholin Gao and Oizheng Qian, 'The Risk and Benefits of Applying Artificial, Intelligence in Business Discussions' (2022) 30 BCP Buusiness & Management 808. doi:10.54691/bcpbm.v30i.2569.

require individuals with specialised skills and knowledge and take place across various fields, including hospitals, offices, and legal consultancies.[9]

AI can enhance contractual negotiations through data analysis, enabling parties to assess market conditions, understand the needs of each party, and identify strengths and weaknesses. It can also evaluate risks, determine optimal negotiation paths, and propose creative solutions. By automating repetitive tasks, AI saves negotiators time, enabling them to focus on more important aspects of the process. However, some scholars contend that AI will not fully replace humans in negotiations, as human input is essential for understanding the other party's position.[10]

Several companies, including IBM, Salesforce, and Alibaba, utilise AI technologies (e.g., Watson, Einstein, and Alibaba DAMO) to assist in negotiations. While AI is increasingly seen as an inevitable force in legal and judicial matters, scholars emphasise that it presents challenges that must be addressed legislatively and technically. They caution that AI should remain under human control to mitigate risks and ensure its ethical integration into society.[11]

4 LEGAL CHALLENGES RELATED TO CONTRACTUAL NEGOTIATION THROUGH AI TECHNOLOGIES

The widespread adoption of artificial intelligence has generated significant legal challenges, particularly with respect to its inherent risks, as well as broader concerns in the domains of research and innovation. This tension largely stems from the accelerating pace of technological advancement, which often outstrips the ability of legal frameworks to adapt, thereby exacerbating these challenges. In this context, the present study focuses on civil liability for damages arising from errors committed by artificial intelligence systems during the contractual negotiation process.

4.1. Civil Liability for Damages Caused by Artificial Intelligence

The issue of civil liability arising from AI errors in negotiations is a recent and complex matter that requires careful consideration from both legal and technical perspectives.

In general, civil liability is defined as the obligation under which a person is liable for remedying the damage incurred by another person due to the acts performed by the

9 Michelle Vaccaro and others, 'Advancing AI Negotiations: New Theory and Evidence from a Large-Scale Autonomous Negotiations Competition' (*arXiv*, 7 July 2025) arXiv:2503.06416v2. doi:10.48550/arXiv.2503.06416.

10 Horst Eidenmüller, 'The Advent of the AI Negotiator: Negotiation Dynamics in the Age of Smart Algorithms' (2025) 20(1) Journal of Business & Technology Law 1.

11 Yousef Abuzir, 'Artificial Intelligence in Legal Practice: Applications, Challenges, and Future Prospects' (2025) 8(1) Journal of Business in the Digital Age 33. doi:10.46238/jobda.1629307.

former, their subordinates, or things for which the former is liable.[12] It can also be defined as "the person's obligation to compensate the damage he caused to another person because of violating an obligation represented in infringing the victim or third parties in whatsoever manner." [13]

In the context of AI-related damages, civil liability refers to the liability of AI for compensating the damage incurred by the victim or a third party as a result of the operation or decision-making of an AI system.[14]

With the growing use of AI in fields like healthcare, civil liability for AI-related errors has become increasingly important—for example, when a robot causes harm to a patient. In such cases, the responsible party must compensate for financial and moral damages, regardless of fault. This strict liability principle applies to institutions using AI, such as hospitals, and serves as a warning to manufacturers of potentially dangerous technologies.

Since the mid-2010s, the European Parliament has shown heightened interest in civil liability arising from AI applications, particularly those involving robots and autonomous vehicles. On 16 February 2017, the Parliament adopted a resolution on Civil Law Rules on Robotics, calling for the development of new legal frameworks that account for the difficulty of proving software errors and the challenges arising from system autonomy in decision-making.[15] This was followed by another resolution on 20 October 2020 (2020/2014(INL)), which recommended the establishment of a comprehensive civil liability system and a balanced compensation scheme capable of addressing damages resulting from the use of AI technologies, recognising the inadequacy of traditional rules based on proving fault and causality.[16]

Member States have responded to these recommendations in various ways. In 2021, Germany adopted a specific law on autonomous driving, requiring the presence of a "technical supervisor" (*technische Aufsicht*) in vehicles and mandated the installation of a "black box"-like device to record driving data for determining liability in the event of an accident. This law represents a practical national model for addressing the evidentiary

12 Ahmed Abu Al-Saud, *The Insurance Policy between Theory and Practice: A Comprehensive Analytical Study* (Dar Al-Fikr Al-Jami'i 2009) [in Arabic].

13 Muhammad Abd al-Zahir Hussein, *The Injured Party's Mistake and Its Impact on Liability* (Dar Al-Nahda Al-Arabiya 2007) [in Arabic].

14 Nikos Th Nikolinakos, *Adapting the EU Civil Liability Regime to the Digital Age: Artificial Intelligence, Robotics, and Other Emerging Technologies* (Law, Governance and Technology Series vol 68, Springer 2024) 377. doi:10.1007/978-3-031-67969-8_8.

15 European Parliament Resolution 2015/2103(INL) of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics [2018] OJ C 252/239 <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=oj:JOC_2018_252_R_0026> accessed 20 April 2025.

16 European Parliament Resolution 2020/2014(INL) of 20 October 2020 with recommendations to the Commission on a Civil Liability Regime for Artificial Intelligence [2021] OJ C 404/107 <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=oj:JOC_2021_404_R_0006> accessed 20 April 2025.

challenges in AI-related incidents, complementing European discussions on the need for a unified framework.[17]

In a broader context, the European Union adopted the Artificial Intelligence Act (AI Act) in July 2024, marking the first binding horizontal European legislation regulating AI technologies. This law is based on a risk-classification methodology, prohibiting systems with "unacceptable risk" and subjecting high-risk systems to stringent technical and procedural obligations, including risk management, mandatory human oversight, maintaining transparent operational records, and reporting malfunctions. The Act also provided for the establishment of national regulatory authorities to supervise compliance, while imposing limited transparency requirements for low-risk systems.[18]

Alongside these developments, the EU Data Act, enacted in 2023, represent a parallel regulatory step aimed at ensuring access to and defining usage rights for data generated by connected devices, thereby providing a supportive legal environment for AI development by facilitating industrial data sharing.

Regarding compensation and liability, the European Commission proposed the AI Liability Directive in 2022, aiming to alleviate the burden of proof for victims by introducing mechanisms such as legal presumptions of causality. However, this proposal faced broad political and legislative disagreements, leading to its withdrawal from the Commission's work programme in 2025, leaving a partial legislative gap in compensation for AI-related damages.[19]

These European developments reflect the European legislator's awareness of the complex nature of AI and the need for innovative legal tools that strike a balance between the requirements of technological innovation and the protection of fundamental rights and contractual interests. Despite such progress, Arab experiences remain more focused on formulating national strategies and general policy frameworks rather than building detailed legislative systems.

In Egypt, the government launched the *National Strategy for Artificial Intelligence* and enacted the Personal Data Protection Law in 2020; however, issues of civil liability for AI damages remain subject to traditional rules in the Civil Code and Consumer Protection Law.[20] In the

17 Johann Laux, Sandra Wachter and Brent Mittelstadt, 'Trustworthy Artificial Intelligence and the European Union AI Act: On the Conflation of Trustworthiness and Acceptability of Risk' (2023) 18(1) Regulation & Governance 3. doi:10.1111/rego.12512.

18 Nuno Sousa e Silva, 'The Artificial Intelligence Act: Critical Overview' (*arXiv*, 30 August 2024) arXiv:2409.00264. doi:10.48550/arXiv.2409.00264.

19 Timo Minssen and others, 'Governing AI in the European Union: Emerging Infrastructures and Regulatory Ecosystems in Health' in Barry Solaiman and I Glenn Cohen (edn), *Research Handbook on Health, AI and the Law* (Edward Elgar 2024) 311. doi:10.4337/9781802205657.ch18.

20 Maha Ramadan Muhammad Battikh, "Civil Liability for Damages Caused by Artificial Intelligence Systems: A Comparative Analytical Study' (2021) 9(5) Legal Journal (Faculty of Law, Cairo University, Khartoum Branch) 1513. doi:10.21608/jlaw.2021.190692 [in Arabic].

UAE, the UAE Strategy for Artificial Intelligence 2031 was launched, and Federal Law No. (45) of 2021 on the Protection of Personal Data was issued, while legislative policy relies on regulatory sandboxes to test systems before widespread deployment, without a specific law yet regulating civil liability for AI-related damages.[21]

An analytical review of the UAE and Egyptian legislative frameworks reveals several fundamental gaps that hinder their ability to address the legal challenges arising from contractual negotiations conducted through AI systems. Chief among these are the absence of explicit legal provisions recognising the legal status of AI systems or robots, and the lack of a precise statutory definition delineating their rights and obligations within contractual frameworks.[22]

Moreover, both jurisdictions lack a dedicated regulatory framework or specialised oversight body responsible for supervising the performance of AI agents or ensuring their compliance with civil liability rules during negotiations. This stands in contrast to the European experience, which has pioneered practical models for establishing independent supervisory entities and linking the use of AI to clear legal duties concerning transparency and accountability.

In light of these findings, the study recommends that UAE and Egyptian lawmakers undertake the following steps:

1. Adopt clear and specific legal definitions of intelligent systems and delineate their contractual use within the Civil Transactions Law.
2. Develop dedicated legal mechanisms to impose liability on AI developers and users for harm resulting from automated actions.
3. Establish specialised regulatory bodies—or dedicated units within existing authorities—to oversee smart contracts and assess the legal performance of the algorithms employed.

The study also recommends the gradual implementation of these mechanisms, drawing on recent European models, to ensure a safe transition toward an AI-driven contractual environment while minimising the regulatory gap.

21 Essam M El Gohary, Ghada El Shabrawy and Sahar Hassib, 'Assessment of the Artificial Intelligence Strategies Announced in the Arab Countries' (2023) 31(3) Egyptian Journal of Development and Planning 1. doi:10.21608/inp.2023.326507 [in Arabic].

22 Adel Salem AlLouzi, Karima KRIM and Mohammad Abdalhafid AlKhamaiseh, 'The Role of Artificial Intelligence and Emerging Technologies in UAE Commercial Transactions Law' (2023) 5(4) Research Journal in Advanced Humanities 156. doi:10.58256/4w202n53.

4.2. Civil Types of Damages Resulting from AI Errors in Contract Negotiations

This review is supported by Chopard and Musy, who argue that AI systems are being increasingly used to aid in the diagnosis and treatment of diseases, thereby reducing the risk of medical errors. They note that such systems also influence the determination and allocation of compensation among doctors and producers of AI systems in cases where patients suffer harm.[23] Conversely, other scholars contend that determining liability in AI-related defects could be complex due to the involvement of multiple stakeholders. Responsibility becomes difficult to establish when defects arise from different sources, such as defective training data, algorithmic biases, or inappropriate system design.[24]

When AI is employed in contract negotiation or dispute resolution, its errors can result in significant financial or legal harm—such as costly decisions, broken agreements, project delays, or unexpected expenses. These mistakes may lead to contractual breaches or tort liability arising from the AI's actions.[25]

1. Operational Damages: Errors by AI systems during negotiations could result in delays of projects and business transactions, leading to additional costs and lost opportunities. This results in weakening the company's competitive position and negatively affecting its reputation.[26]

2. Legal Damages: In addition to operational damages, AI errors in negotiation can result in serious legal problems, including contractual violations or torts (harmful acts), which may lead to expensive judicial disputes.

3. Financial Damages: When AI errors occur in negotiating contracts, this can result in substantial financial losses to the negotiating parties, which could include concluding unprofitable or unfavourable agreements, missing opportunities for profit and business growth.

23 Bertrand Chopard and Olivier Musy, 'Market For Artificial Intelligence in Health Care and Compensation for Medical Errors' (2023) 75 International Review of Law and Economics 106153. doi:10.1016/j.irle.2023.106153.

24 Miriam C Buiten, 'Product Liability for Defective AI' (2024) 57(1) European Journal of Law and Economics 239. doi:10.1007/s10657-024-09794-z.

25 Abdelrazek Wahba Sayedahmed, 'Civil Liability for Artificial Intelligence Damages: An Analytical Study' (2020) 43 Generation of In-depth Legal Research Journal 11 [in Arabic].

26 Ahmed M Al-Hawamdeh and Tariq K Alhasan, 'Smart Robots and Civil Liability in Jordan: A Quest for Legal Synthesis in the Age of Automation' (2024) 16(2) Jordanian Journal of Law and Political Science 52. doi:10.35682/jjlps.v16i2.743.

Civil liability for AI errors in negotiations is established on a group of main legal pillars, foremost among them being the principle of tort (harmful act), which requires compensation for damage resulting from a harmful act or negligence by the entity responsible for the AI.[27]

This liability is grounded in the principles of contractual liability, under which the entity in charge of AI is obliged to exert due diligence to prevent such damages. In the event of violating such an obligation, it shall compensate the victims for the losses they incur or the damages resulting from the AI errors during negotiations.[28]

4.3. Conditions for Civil Liability for Errors Committed by Artificial Intelligence during Negotiations

Civil liability for errors committed by AI during negotiations arises upon the fulfilment of the following condition:[29]

1. Incurrence of Actual Damage: Civil liability cannot be established unless the AI error results in actual damage to the negotiating parties, whether financial or moral. The damage must be direct and causally linked to the AI's error.[30]
2. Commission of an AI Error: Civil liability arises when the AI system commits an error while performing its negotiation functions. The error must stem from deficiencies in programming or AI performance, rather than from events constituting force majeure or other factors beyond the control of the entity in charge.[31]
3. Causation Relationship: There must be a causation relationship between the AI error and the damage incurred by the negotiating parties. In other words, the damage

27 Bashar Talal Momani and others, 'Securing Privacy: Safeguarding against Cyber Threats in the UAE and Morocco' (2024) 5(3) Global Privacy Law Review 126. doi:10.54648/gplr2024018; Mohammed Al Morsi Zahra, *Non-administrative sources of obligation in the Civil Transactions Law of the United Arab Emirates: Harmful Acts and Beneficial Acts* (UAE University Press 2003) [in Arabic]; Nasser Mohammed Abdullah Sultan, *Liability for the Act of Things Requiring Special Care and Mechanical Machinery in Light of the UAE Civil Code Compared to the Egyptian Civil Code* (Al-Halabi Legal Publications 2005) [in Arabic]; Osama Ahmed Badr, *The Concept of Guardianship (Hirasah) in Civil Liability: A Comparative Study* (Dar Al-Nahda Al-Arabiya 2004) [in Arabic].

28 Ibrahim Al-Desouki Abullail, 'Smart Contracts and Artificial Intelligence and their Role in the Automation of Contracts and Legal Acts: Study of the Role of Scientific Progress in the Development of Contract Theory' (2020) 44(4/1) Journal of Law. doi:10.34120/jol.v44i4.2545 [in Arabic].

29 Abdel Razzaq Ahmed Al-Sanhouri, *The Intermediary in Explaining the New Civil Law: The Theory of Obligation in General*, pt 3 (Manshaet Al-Ma'arif 2004) [in Arabic].

30 Martin Ebers, 'Liability for Artificial Intelligence and EU Consumer Law' (2021) 12(2) JIPITEC 204.

31 Reza Farajpour, 'The Role of Civil Liability in Artificial Intelligence Laws from the Perspective of Major Global Legal Systems' (2025) 5(2) Journal of Law and Political Studies 182. doi:10.48309/jlps.2025.518711.1353.

must be the direct and inevitable outcome of the AI error in negotiation and may not be the result of any other factors.[32]

In defining the party civilly liable for compensating damages resulting from the use of AI in contractual negotiations, liability may fall upon one or more of the following:[33]

1. Liability of the Robot Manufacturer and Programmer: The manufacturer, developer, or programmer of an AI system may incur civil liability where damage arises from defects in design, errors in programming, or negligence in the manufacturing or development process. In such cases, liability is generally governed by product liability laws, which impose a legal obligation on manufacturers to ensure that their products are safe, effective, and free from defects.[34]

2. Liability for Use (End-User Responsibility): Where harm results from improper use of the AI system—such as failure to adhere to usage guidelines or intentional misuse—the end user may be held liable. In such instances, liability is determined in accordance with the general principles governing fault-based liability for misuse or negligence.[35]

3. Contracts and Agreements: Contracts governing the use of AI usually include provisions determining the liability for the damages resulting from the AI errors. Such provisions may release the manufacturer from liability in some cases or limit the amount of compensation. Therefore, such contracts must be carefully reviewed to determine the liable party in case of error that results in damage to third parties or to determine the manner of dividing the liability among the different parties.[36]

4. Training and Maintenance: Liability may be borne by the user if the error is the result of the lack of training or lack of periodical maintenance of the AI.

5. Determining the Error: In case of error, it is necessary to determine whether the cause lies in programming, a technical error, or user misuse. This requires a technical investigation, involving experts in technology and software.

32 Gabriele Buchholtz, 'Artificial Intelligence and Legal Tech: Challenges to the Rule of Law' in Thomas Wischmeyer and Timo Rademacher (eds), *Regulating Artificial Intelligence* (Springer 2020) 175. doi:10.1007/978-3-030-32361-5_8.

33 Khaled Abdel Fattah Saqr, *Rules and Provisions of Criminal and Civil Liability for Architects, Doctors, Contractors, Property Owners, and Custodians* (Dar Mahmoud for Publishing and Distribution 2024) [in Arabic].

34 Alice Guerra, Francesco Parisi and Daniel Pi, 'Liability for Robots I: Legal Challenges' (2022) 18(3) Journal of Institutional Economics 331. doi:10.1017/S1744137421000825.

35 Philipp Hacker, 'The European AI Liability Directives: Critique of a Half-Hearted Approach and Lessons for the Future' (2023) 51 Computer Law & Security Review 105871. doi:10.1016/j.clsr.2023.105871.

36 Hannes Claes and Maarten Herbosch, 'Artificial Intelligence and Contractual Liability Limitations: A Natural Combination?' (2023) 31(2/3) European Review of Private Law 469. doi:10.54648/erpl2023027.

6. Insurance: With the rapid progress in the field of AI and robotics, legislation can develop to include specific provisions related to the liability for AI errors. For example, the expected laws can consist of existing or providing special insurance to cover the damages resulting from the use of AI, which could add a protection layer for the users. Moreover, AI has a significant impact on the insurance industry and poses a future challenge in light of the potential errors that could occur.[37]

In general, determining civil liability in this context is dependent on identifying the primary cause of the error and the manner in which local laws deal with such lawsuits. Given the complexity of these issues, it is also necessary to consult specialised legal experts to obtain accurate and context-specific advice regarding liability for AI-related damages.

4.4. Victim's Obligations in Civil Claims Arising from AI Errors in Negotiation

In civil litigation concerning AI-related errors during contractual negotiations, the claimant is subject to specific procedural and evidentiary obligations.

First, the victim must comply with strict procedural time limits for initiating legal action, as failure to observe statutory deadlines often results in the forfeiture of the right to compensation.[38]

Second, the burden of proof rests on the claimant, who must demonstrate both the existence of an AI malfunction and establish a direct causal nexus between the system's error and the harm incurred. Given the autonomous and opaque nature of AI decision-making, this requirement represents a significant legal challenge.

Third, the claimant is required to provide sufficient evidence of the alleged harm, which may include financial statements, contractual records, and other forms of documentary evidence. Furthermore, technical documentation and expert testimony are often essential to establish whether the damage was attributable to a programming deficiency, a system malfunction, or improper human use.

Recent scholarship has increasingly emphasised the necessity of procedural innovations, such as evidentiary presumptions and reversed burdens of proof, to effectively balance victims' rights with the complexities of AI accountability frameworks in the European context.[39]

37 Chris Lamberton, Damiano Brigo and Dave Hoy, "Impact of Robotics, RPA and AI on the Insurance Industry: Challenges and Opportunities' (2017) 4(1) Journal of Financial Perspectives 8.

38 Ana Taveira da Fonseca, Elsa Vaz de Sequeira and Luís Barreto Xavier, 'Liability for AI-Driven Systems' in Henrique Sousa Antunes and others (eds), *Multidisciplinary Perspectives on Artificial Intelligence and the Law* (Law, Governance and Technology Series, Springer 2023) 299. doi:10.1007/978-3-031-41264-6_16.

39 Ebers (n 30).

The methods by which parties may deny civil liability for AI-related errors in negotiation have given rise to several legal questions, including whether the AI user can rely on the foreign cause.[40]

The concept of foreign cause encompasses the urgent or sudden accidents, force majeure events, acts of third parties, acts of the victim, or technical incidents such as breakdowns and viruses affecting AI systems. Under certain conditions, the AI user may rely on such arguments to deny liability. This possibility finds legal support in Article 373 of the Egyptian Civil Code[41] and Article 287 of the UAE Federal Civil Code.[42]

The European Parliament (EP) has paid particular attention to the issue of civil liability for damages caused by AI systems, including AI software embedded in robots and autonomous driving cars. On 17 February 2017, the EP adopted a series of recommendations related to the civil liability for the damages incurred by third parties. [43]

These recommendations highlighted two major challenges:

1. The difficulty of attributing error to AI systems under traditional civil liability frameworks, which typically require the establishment of human fault or negligence as a precondition for civil liability.

2. The limitations of holding AI software liable for cases in which AI can make independent and subjective decisions. In such situations, it becomes problematic to identify a "defect" that caused such damage and the causal link between the assumed defect and the resulting damage.[44]

Accordingly, the European Parliament concluded that the general rules of civil liability are insufficient for addressing the damages caused by AI software and applications. It also urged enacting a special legal framework to accommodate and regulate them in proportion to the nature of AI applications.[45]

40 Muhammad Labib Shanab, *Responsibility for Things: A Comparative Study* (2nd edn, Al-Wafa Legal Library 2009) [in Arabic].

41 Law of the Arab Republic of Egypt No 131 of 1948 'Civil Code' (amended 20 July 2025) <https://eg.andersen.com/translation-law-131-1948/> accessed 25 July 2025. Article 373 of the Egyptian Civil Code stipulates, "The obligor shall be released from liability if the obligor proves that honoring the liability has become impossible for a foreign cause beyond the obligor's control."

42 Federal Law of the United Arab Emirates No 5 of 1985 'On the Civil Transactions Law of the United Arab Emirates' (Civil Code) (amended 27 September 2020) <https://www.lexismiddleeast.com/law/UnitedArabEmirates/Law_5_1985> accessed 20 April 2025. Article 287 of the UAE Civil Code stipulates, "If a person proves that the damage was due to a foreign cause beyond his control, such as the acts of god, sudden accident, force majeure event, third party's act or the victim's acts, the person shall not be liable for the guarantee unless law or the agreement stipulates otherwise."

43 European Parliament Resolution 2015/2103(INL) (n 15).

44 This difficulty is because some artificial intelligence programs can self-learn from their own changing experiences, which enable them to interact in the external environment in a unique and unexpected way.

45 Nasr Aboul Fotouh Farid Hassan, 'Smart Contracts between Reality and Prospects: An Analytical Study' (2020) 28(2) Journal of Security and Law 499. doi:10.54000/0576-028-002-009 [in Arabic].

Among the legislative responses to this call, Germany's 2017 amendment to its Road Traffic Act stands out. This legislation introduced specific rules governing the civil liability of autonomous cars, including the following provisions:

1. The driver must be present in the vehicle at all times while it is in motion.
2. The driver must retain control of the vehicle when the AI system prompts manual intervention, particularly when the system requires that the driver take over the steering wheel.
3. Every autonomous vehicle must be equipped with a black box, similar to those in aeroplanes, to record specific data, including the vehicle's itinerary and the driver's control status at the time of an incident—specifically, whether the accident occurred while the vehicle was under manual control or autonomous operation. If the accident occurs while the vehicle is operating during autonomous mode, the liability shall be borne by the car manufacturer. However, if the accident occurs due to the driver's failure, for instance, to take control despite receiving notifications and warnings from the AI system, the driver shall bear the liability.[46]

4.5. Critical Reflections on Regulatory Complexity and Comparative Prospects

Amid the rapid evolution of AI and its growing role in automated contractual negotiations, the European Law Institute (ELI) issued its 2022 Principles on AI, emphasising the protection of the right to human review of automated decisions and the necessity of preventing a denial of access to justice arising from reliance on intelligent negotiation systems.[47] These guidelines further advocate for the modernisation of traditional legal categories, particularly the notion of "product" to include intelligent software, thereby aligning with evolving approaches to product liability in light of increasing automation.[48]

While these principles provide a valuable theoretical foundation for AI governance within European private law, their scope remains largely Eurocentric and insufficiently tailored to the specificities of non-Western jurisdictions. This underscores the need for comparative research, particularly within Arab legal systems, to assess the adaptability of these principles in light of domestic legal traditions, regulatory frameworks, and socio-cultural constraints. Recent scholarship stresses that legal responses to AI must avoid a mere transplantation of European models, and instead develop context-sensitive frameworks capable of addressing local needs while engaging with global standards of AI governance.[49]

46 Nasr Farid Hassan, 'Some Legal Aspects Related to the Operation of Self-Driving Vehicles According to Dubai Law No. (9) of 2023' (2024) 21(4) University of Sharjah Journal of Legal Sciences. doi:10.36394/jls.v21.i4.10 [in Arabic].

47 Rodriguez de las Heras Ballell (n 1) 21-2.

48 ibid 12-3.

49 Jānis Kārkliņš, 'Artificial Intelligence and Civil Liability' (2020) 13 Journal of the University of Latvia: Law 164. doi:10.22364/jull.13.10.

This research trajectory represents a crucial first step toward establishing an Arab perspective on AI-related liability in contractual negotiations, laying the groundwork for a comparative legal framework that balances technological innovation with the protection of fundamental rights.[50]

4.6. Distinguishing Contractual and Tortious Liability in AI-Related Cases

In AI-related disputes—particularly within contractual negotiation contexts—the distinction between contractual and tortious liability is of fundamental importance due to the technical and operational complexity of intelligent systems. **Contractual liability** arises where a contractual relationship exists between the user or injured party and the developer or operator, entailing obligations such as performance or product safety. In contrast, **tortious liability** applies in the absence of such a relationship and may be based on fault, negligence, or, in some cases, strict liability—especially with high-risk AI systems.

The **European Law Institute** emphasises that traditional legal frameworks are no longer sufficient and calls for extending tortious liability to cover harm caused by high-risk AI, even in the absence of a contractual relationship.[51] Similarly, **Cogen et al.** contend that tortious liability in this context necessitates a reconsideration of fault and evidentiary standards—potentially shifting the burden of proof or introducing legal presumptions to facilitate claims.[52]

Recent studies[53] highlight the practical overlap between contractual and tortious liability, particularly in smart or long-term contracts involving both human and automated elements. This overlap necessitates a redefinition of the conceptual and legislative boundary between the two regimes.

Effective legal regulation of civil liability for AI errors cannot rely solely on classical doctrines. Instead, a **hybrid legal framework** is needed—one that accommodates the unique characteristics of intelligent systems while ensuring meaningful protection for both contracting parties and third parties.

50 Esther Salmerón-Manzano, 'Emerging Technologies, Law and Policies' (2025) 14 Laws 28. doi:10.3390/laws14020028.

51 Rodriguez de las Heras Ballell (n 1) 11-2.

52 Orian Dheu and Jan De Bruyne, 'Artificial Intelligence and Tort Law: A 'Multi-faceted' Reality' (2023) 31(2/3) European Review of Private Law 261. doi:10.54648/erpl2023021.

53 Sharmila Ramachandaran and others, 'Exploring the Challenges of AI-driven Business Intelligence Systems in the Malaysian Insurance Industry' (F1000Research, 22 April 2025). doi:10.12688/f1000research.163354.1.

4.7. Scope of Damage Covered Under Product Liability

In the context of applying product liability rules within the European Union to AI technologies—particularly in contractual negotiations—the delineation of compensable harm emerges as a pivotal issue for ensuring a balance between adequate protection of victims and avoiding disproportionate legal burdens on producers. The recent reform introduced by Directive (EU) 2024/2853 on liability for defective products represents a fundamental update, specifying in an exhaustive manner the categories of compensable damage: death or personal injury, including medically recognised psychological harm; damage to property owned by natural persons (excluding the defective product itself and property used exclusively for professional purposes); and destruction or corruption of data, provided that the data is not used for professional purposes.[54]

The Directive further clarifies that "pure economic loss," as well as harms linked to privacy violations or discrimination, do not in themselves give rise to liability under this framework, although such harms may be addressed under other liability regimes at the national level. This relatively narrow definition reflects the European approach of facilitating effective redress for individuals affected by defective digital products and software—including AI systems—without transforming product liability into a catch-all mechanism for compensating every form of immaterial or purely economic loss.[55]

From a comparative perspective, recent legal scholarship underscores that the reform of product liability rules was driven by the increasing complexity of digital products, supply chains, and the integration of software and machine learning components, while maintaining the logic of strict liability for producers. The scope of compensable damage was deliberately circumscribed to preserve legal certainty and prevent "liability inflation" that could deter innovation. At the same time, evidentiary burdens have been relaxed in favour of claimants, introducing presumptions of defect and causation to mitigate the technical difficulties of proving harm in cases involving AI technologies.

By contrast, Arab civil law systems adopt a broader approach. In Egyptian law, tort liability is founded on the elements of fault, harm, and causation,[56] with wide recognition of compensation for both material and moral damages, without the strict categorical limitations found in the EU framework. Similarly, the UAE Civil Transactions Law[57] explicitly provides for compensation of both material and moral

54 Claudio Novelli and others, 'Generative AI in EU law: Liability, privacy, intellectual property, and cybersecurity' (2024) 55 Computer Law & Security Review 106066. doi:10.1016/j.clsr.2024.106066.

55 Beatriz Botero Arcila, 'AI Liability in Europe: How Does it Complement Risk Regulation and Deal with the Problem of Human Oversight?' (2024) 54 Computer Law & Security Review 106012. doi:10.1016/j.clsr.2024.106012.

56 Law of the Arab Republic of Egypt No 131 of 1948 (n 41) art 163 et seq.

57 Federal Law of the United Arab Emirates No 5 of 1985 (n 42).

harm, with Article 293 expressly recognising moral damages and extending compensation to the victim's heirs in specific cases.[58]

This structure results in a broader remedial scope in Egypt and the UAE than under the European product liability regime, encompassing moral harm and, in practice, certain forms of economic loss, albeit subject to judicial interpretation and doctrinal limitations. Consequently, the transposition of EU product liability rules into Arab jurisdictions requires caution, as the substantive scope of compensable harm and the underlying policy objectives differ significantly: while the European framework is narrowly tailored to protect natural persons under a specialised product liability regime, Arab civil law systems operate within general liability frameworks that are more expansive in their remedial reach.[59]

5 CALCULATING THE COMPENSATION FOR THE DAMAGES RESULTING FROM THE AI ERRORS IN NEGOTIATION

When a defendant fails to rebut allegations of liability for harm caused by an AI system during contractual negotiations, courts will ordinarily order the defendant to pay compensation commensurate with the loss sustained by the claimant. Assessing damages in such cases requires a careful and multifaceted exercise.

First, courts must identify and quantify direct and indirect economic losses, including lost profits (lucrum cessans), additional costs reasonably incurred by the injured party to mitigate or remedy the harm, and losses arising from frustrated or rescinded contracts that resulted from the AI malfunction.

Second, non-pecuniary harms—commonly described as moral damages—must be examined where relevant; such harms may encompass reputational injury, loss of goodwill, and the adverse commercial consequences arising from client attrition or the collapse of strategic relationships attributable to the AI failure.

Third, the evidentiary process necessarily demands both conventional documentary proof (such as financial records, contracts, correspondence) and technical proof, including system logs, incident reports, forensic analyses, and expert testimony on software behaviour and fault.[60]

58 Pierre Mallet and Hala Nassar, 'Consensual Terms Modifying Contractual Liability in the Light of UAE Law: A Comparative Study with French Law' (2024) 7(4) Access to Justice in Eastern Europe 218. doi:10.33327/AJEE-18-7.4-a000107.

59 Sarah Zein, 'The Civil Liability for Artificial Intelligence' (2023) 2022(1) BAU Journal of Legal Studies 14. doi:10.54729/2958-4884.1110.

60 W Nicholson Price II, Sara Gerke and I Glenn Cohen, 'Liability for Use of Artificial Intelligence in Medicine' in Barry Solaiman and I Glenn Cohen (eds), *Research Handbook on Health, AI and the Law* (Edward Elgar Publishing 2024) 150. doi:10.4337/9781802205657.ch09.

Insurance plays a dual role in this ecosystem. On one hand, civil-liability insurance—whether as bespoke AI performance policies or as extensions of existing technology-E&O/cyber covers—can provide direct compensatory relief to victims and spread residual risk across underwriters,[61] thereby reducing the immediate financial exposure of developers and deployers. On the other hand, well-designed insurance markets can foster responsible innovation by incentivising appropriate governance, testing, and maintenance practices; insurers may require conformity with best practices as underwriting conditions.[62]

Nevertheless, insurance solutions have limitations: apportioning liability among developers, vendors, and end-users is often technically and contractually complex; the scarcity of historical loss data for novel AI failure modes complicates underwriting and pricing; and rapid technological change risks producing coverage gaps unless policy wordings and regulatory guidance evolve in tandem with technology.[63]

From a legal-policy standpoint, improving victim protection in AI negotiation scenarios requires a three-pronged approach:

1. Clearer substantive liability rules, including calibrated rules on causation and presumptions where appropriate.
2. Robust evidentiary and technical infrastructures for incident analysis and attribution.
3. Development of insurance mechanisms and regulatory incentives that both compensate victims and promote risk-reducing behaviour by market participants.[64]

Under Egyptian law, tort liability rests on the traditional triad of fault, damage, and causal link,[65] with courts routinely recognising both material and moral damages (including loss of reputation and consequential commercial losses), provided they are proven, foreseeable, and proximate to the wrongful act. Consequently, Egyptian practice generally permits recovery for lost profits and reputational harm when such losses can be substantiated and causally linked to the defendant's conduct or the malfunctioning system.[66]

The UAE legal framework similarly follows a general tort model,[67] recognising compensation for both material and moral injury (see Article 293 regarding moral

61 Al-Saud (n 12).

62 Elisa Luciano, Matteo Cattaneo and Ron Kenett, 'Adversarial AI in Insurance:Pervasiveness and Resilience' (*arXiv*, 17 January 2023) arXiv:2301.0752015. doi:10.48550/arXiv.2301.07520.

63 'Assurance RC Pro adaptée à l'IA generative, les clauses indispensables en 2025' (*Hiscox le blog*, 3 April 2025) <https://www.hiscox.fr/blog/assurance-rc-pro-adaptee-lia-generative-les-clauses-indispensables-en-2025> accessed 20 April 2025.

64 Ramachandaran and others (n 53).

65 Law of the Arab Republic of Egypt No 131 of 1948 (n 41) arts 163 ff.

66 Mohammad Ahmed Abdeen, *Compensation between Material and Moral Damage* (Mansha'at Al-Ma'arif 2002) [in Arabic]; Mohsen Abdel Hamid Ibrahim Al-Bayeh, *The General Theory of Obligations: Involuntary Sources*, pt 2 (2nd edn, Dar Al Nahda Al Arabiya 2011) [in Arabic].

67 Federal Law of the United Arab Emirates No 5 of 1985 (n 42).

damages). Thus, both jurisdictions operate within broad remedial systems that—at least doctrinally—allow compensation for economic and non-economic harms arising from AI failures, subject to the usual constraints of proof, foreseeability, and causation. These features contrast with the evolving European product-liability approach, which narrowly defines compensable damage under the product liability instrument while relying on other regimes for purely economic or privacy-related harms. Accordingly, transplanting European product-liability rules into Egyptian or Emirati legal contexts would require careful calibration to account for the more expansive remedial traditions and evidentiary practices in those jurisdictions.[68]

6 FINDINGS

6.1. Results

The analysis reveals that the integration of AI into contractual negotiations presents significant legal challenges, primarily due to the absence of explicit legislative regulation in Arab jurisdictions and the lack of legal recognition for the autonomous features of AI systems. Egyptian and Emirati civil codes continue to rely on traditional liability structures—built on fault, harm, and causation—without providing tailored provisions for AI-driven decision-making. This results in regulatory uncertainty concerning the allocation of liability among developers, operators, and end-users.

A comparative examination of the European Union reveals a more advanced, though still evolving, framework. Instruments such as the AI Act (Regulation (EU) 2024/1689), the Data Act (Regulation (EU) 2023/2854), and the ongoing debate around the AI Liability Directive collectively seek to mitigate evidentiary burdens on victims, impose obligations of transparency and human oversight, and narrowly define compensable damages under product liability rules. The EU approach strikes a balance between protecting victims and safeguarding innovation, introducing rebuttable presumptions of defect and causation to address the opacity of AI systems.

By contrast, Egyptian and Emirati legal systems adopt broader remedial traditions. Both jurisdictions allow compensation for material and moral damages, including reputational harm and lost profits, without the restrictive categories found in the EU framework. However, the absence of statutory definitions of "intelligent systems," the lack of dedicated supervisory bodies, and the reliance on general civil code provisions hinder their ability to address the unique challenges posed by AI errors in negotiation.

68 Bakhit Muhammad Al-Daja, *Artificial Intelligence: Challenges of Contemporary Civil Liability* (Dar Al-Thaqafa for Publishing and Distribution 2023) [in Arabic].

Consequently, the comparative analysis underscores three key findings:

1. Regulatory Gap: Arab jurisdictions lack specialised legislation to address AI liability, in contrast to the EU's incremental regulatory reforms.
2. Doctrinal Tension: The strict liability logic of European product law diverges from the broader remedial approach in Arab civil law, complicating any direct transplantation of rules.
3. Evidentiary Complexity: Across both systems, proving causation and fault in AI-related harm remains a fundamental obstacle, necessitating novel legal and technical mechanisms.

6.2. Recommendations

Based on these findings, the study proposes several normative measures:

1. Adopt precise legal definitions of AI and intelligent systems within civil codes, ensuring clarity in determining rights, duties, and liability.
2. Establish specialised regulatory or supervisory authorities in Arab jurisdictions to oversee the use of AI in contractual contexts, drawing inspiration from the EU model of independent oversight bodies.
3. Develop hybrid liability frameworks that integrate contractual and tortious doctrines with calibrated presumptions of defect and causation, thereby easing the burden of proof for victims while maintaining fairness to developers and users.
4. Strengthen insurance mechanisms tailored to AI-related risks, both as compensatory instruments and as tools for incentivising responsible AI governance and risk management practices.
5. Ensure context-sensitive legal transposition by avoiding wholesale adoption of European models and instead designing frameworks that respect Arab legal traditions while engaging with global standards of AI regulation.

Together, these recommendations aim to bridge the current regulatory gap, enhance legal certainty, and promote a balanced framework that simultaneously safeguards victims, supports innovation, and ensures accountability in AI-driven contractual negotiations.

REFERENCES

1. Abdeen MA, *Compensation between Material and Moral Damage* (Mansha'at Al-Ma'arif 2002) [in Arabic]
2. Abullail IAD, 'Smart Contracts and Artificial Intelligence and their Role in the Automation of Contracts and Legal Acts: Study of the Role of Scientific Progress in the Development of Contract Theory' (2020) 44(4/1) Journal of Law. doi:10.34120/jol.v44i4.2545 [in Arabic]
3. Abuzir Y, 'Artificial Intelligence in Legal Practice: Applications, Challenges, and Future Prospects' (2025) 8(1) Journal of Business in the Digital Age 33. doi:10.46238/jobda.1629307
4. Al-Bayeh MAH, *The General Theory of Obligations: Involuntary Sources*, pt 2 (2nd edn, Dar Al Nahda Al Arabiya 2011) [in Arabic]
5. Al-Daja BM, *Artificial Intelligence: Challenges of Contemporary Civil Liability* (Dar Al-Thaqafa for Publishing and Distribution 2023) [in Arabic]
6. Al-Hawamdeh AM and Alhasan TK, 'Smart Robots and Civil Liability in Jordan: A Quest for Legal Synthesis in the Age of Automation' (2024) 16(2) Jordanian Journal of Law and Political Science 52. doi:10.35682/jjlps.v16i2.743
7. AlLouzi AS, KRIM K and AlKhamaiseh MA, 'The Role of Artificial Intelligence and Emerging Technologies in UAE Commercial Transactions Law' (2023) 5(4) Research Journal in Advanced Humanities 156. doi:10.58256/4w202n53
8. Al-Sanhouri ARA, *The Intermediary in Explaining the New Civil Law: The Theory of Obligation in General*, pt 3 (Manshaet Al-Ma'arif 2004) [in Arabic]
9. Al-Saud AA, *The Insurance Policy between Theory and Practice: A Comprehensive Analytical Study* (Dar Al-Fikr Al-Jami'i 2008) [in Arabic]
10. Badr OA, *The Concept of Guardianship (Hirasah) in Civil Liability: A Comparative Study* (Dar Al-Nahda Al-Arabiya 2004) [in Arabic]
11. Battikh MRM, "Civil Liability for Damages Caused by Artificial Intelligence Systems: A Comparative Analytical Study' (2021) 9(5) Legal Journal (Faculty of Law, Cairo University, Khartoum Branch) 1513. doi:10.21608/jlaw.2021.190692 [in Arabic]
12. Bertolin A, *Artificial Intelligence and Civil Liability: A European Perspective* (Study, European Parliament's Policy Department for Justice 2025)
13. Botero Arcila B, 'AI Liability in Europe: How Does it Complement Risk Regulation and Deal with the Problem of Human Oversight?' (2024) 54 Computer Law & Security Review 106012. doi:10.1016/j.clsr.2024.106012
14. Buchholtz G, 'Artificial Intelligence and Legal Tech: Challenges to the Rule of Law' in Wischmeyer T and Rademacher T (eds), *Regulating Artificial Intelligence* (Springer 2020) 175. doi:10.1007/978-3-030-32361-5_8

15. Buiten MC, 'Product Liability for Defective AI' (2024) 57(1) European Journal of Law and Economics 239. doi:10.1007/s10657-024-09794-z

16. Chopard B and Musy O, 'Market for Artificial Intelligence in Health Care and Compensation for Medical Errors' (2023) 75 International Review of Law and Economics 106153. doi:10.1016/j.irle.2023.106153

17. Claes H and Herbosch M, 'Artificial Intelligence and Contractual Liability Limitations: A Natural Combination?' (2023) 31(2/3) European Review of Private Law 469. doi:10.54648/erpl2023027

18. da Fonseca AT, Vaz de Sequeira E and Barreto Xavier L, 'Liability for AI-Driven Systems' in Sousa Antunes H and others (eds), *Multidisciplinary Perspectives on Artificial Intelligence and the Law* (Law, Governance and Technology Series, Springer 2023) 299. doi:10.1007/978-3-031-41264-6_16

19. Dheu O and De Bruyne J, 'Artificial Intelligence and Tort Law: A 'Multi-faceted' Reality' (2023) 31(2/3) European Review of Private Law 261. doi:10.54648/erpl2023021

20. Ebers M, 'Liability for Artificial Intelligence and EU Consumer Law' (2021) 12(2) JIPITEC 204.

21. Eidenmüller H, 'The Advent of the AI Negotiator: Negotiation Dynamics in the Age of Smart Algorithms' (2025) 20(1) Journal of Business & Technology Law 1

22. El Gohary EM, El Shabrawy G and Hassib S, 'Assessment of the Artificial Intelligence Strategies Announced in the Arab Countries' (2023) 31(3) Egyptian Journal of Development and Planning 1. doi:10.21608/inp.2023.326507 [in Arabic]

23. Farajpour R, 'The Role of Civil Liability in Artificial Intelligence Laws from the Perspective of Major Global Legal Systems' (2025) 5(2) Journal of Law and Political Studies 182. doi:10.48309/jlps.2025.518711.1353

24. Gao Z and Qian O, 'The Risk and Benefits of Applying Artificial, Intelligence in Business Discussions' (2022) 30 BCP Buusiness & Management 808. doi:10.54691/bcpbm.v30i.2569

25. Guerra A, Parisi F and Pi D, 'Liability for Robots I: Legal Challenges' (2022) 18(3) Journal of Institutional Economics 331. doi:10.1017/S1744137421000825

26. Hacker P, 'The European AI Liability Directives: Critique of a Half-Hearted Approach and Lessons for the Future' (2023) 51 Computer Law & Security Review 105871. doi:10.1016/j.clsr.2023.105871

27. Hassan NAF, 'Smart Contracts between Reality and Prospects: An Analytical Study' (2020) 28(2) Journal of Security and Law 499. doi:10.54000/0576-028-002-009 [in Arabic]

28. Hassan NF, 'Some Legal Aspects Related to the Operation of Self-Driving Vehicles According to Dubai Law No. (9) of 2023' (2024) 21(4) University of Sharjah Journal of Legal Sciences. doi:10.36394/jls.v21.i4.10 [in Arabic].

29. Hussein MA, *The Injured Party's Mistake and its Impact on Liability* (Dar Al-Nahda Al-Arabiya 2007) [in Arabic]

30. Kārkliņš J, 'Artificial Intelligence and Civil Liability' (2020) 13 Journal of the University of Latvia: Law 164. doi:10.22364/jull.13.10

31. Lamberton C, Brigo D and Hoy D, "Impact of Robotics, RPA and AI on the Insurance Industry: Challenges and Opportunities' (2017) 4(1) Journal of Financial Perspectives 8.

32. Laux J, Wachter S and Mittelstadt B, 'Trustworthy Artificial Intelligence and the European Union AI Act: On the Conflation of Trustworthiness and Acceptability of Risk' (2023) 18(1) Regulation & Governance 3. doi:10.1111/rego.12512

33. Luciano E, Cattaneo M and Kenett R, 'Adversarial AI in Insurance:Pervasiveness and Resilience' (*arXiv*, 17 January 2023) arXiv:2301.0752015. doi:10.48550/arXiv.2301.07520

34. Mallet P and Nassar H, 'Consensual Terms Modifying Contractual Liability in the Light of UAE Law: A Comparative Study with French Law' (2024) 7(4) Access to Justice in Eastern Europe 218. doi:10.33327/AJEE-18-7.4-a000107

35. Minssen T and others, 'Governing AI in the European Union: Emerging Infrastructures and Regulatory Ecosystems in Health' in Solaiman B and Cohen IG (edn), *Research Handbook on Health, AI and the Law* (Edward Elgar 2024) 311. doi:10.4337/9781802205657.ch18

36. Momani BT and others, 'Securing Privacy: Safeguarding against Cyber Threats in the UAE and Morocco' (2024) 5(3) Global Privacy Law Review 126. doi:10.54648/gplr2024018

37. Nikolinakos NT, *Adapting the EU Civil Liability Regime to the Digital Age: Artificial Intelligence, Robotics, and Other Emerging Technologies* (Law, Governance and Technology Series vol 68, Springer 2024) 377. doi:10.1007/978-3-031-67969-8_8

38. Novelli C and others, 'Generative AI in EU law: Liability, privacy, intellectual property, and cybersecurity' (2024) 55 Computer Law & Security Review 106066. doi:10.1016/j.clsr.2024.106066

39. Price WN II, Gerke S and Cohen IG, 'Liability for Use of Artificial Intelligence in Medicine' in Solaiman B and Cohen IG (eds), *Research Handbook on Health, AI and the Law* (Edward Elgar Publishing 2024) 150. doi:10.4337/9781802205657.ch09

40. Ramachandaran S and others, 'Exploring the Challenges of AI-driven Business Intelligence Systems in the Malaysian Insurance Industry' (*F1000Research*, 22 April 2025). doi:10.12688/f1000research.163354.1.

41. Rodriguez de las Heras Ballell T, *Guiding Principles for Automated Decision-Making in the EU* (ELI Innovation Paper, European Law Institute 2022)

42. Salmerón-Manzano E, 'Emerging Technologies, Law and Policies' (2025) 14 Laws 28. doi:10.3390/laws14020028

43. Saqr KAF, *Rules and Provisions of Criminal and Civil Liability for Architects, Doctors, Contractors, Property Owners, and Custodians* (Dar Mahmoud for Publishing and Distribution 2024) [in Arabic].

44. Sayedahmed AW, 'Civil Liability for Artificial Intelligence Damages: An Analytical Study' (2020) 43 Generation of In-depth Legal Research Journal 11. [in Arabic]

45. Sée A, 'La régulation des algorithmes : un nouveau modèle de globalisation?' (2019) 5 Revue française de droit administratif 830

46. Shanab ML, *Responsibility for Things: A Comparative Study* (2nd edn, Al-Wafa Legal Library 2009) [in Arabic]

47. Sousa e Silva N, 'The Artificial Intelligence Act: Critical Overview' (*arXiv*, 30 August 2024) arXiv:2409.00264. doi:10.48550/arXiv.2409.00264

48. Sultan NMA, *Liability for the Act of Things Requiring Special Care and Mechanical Machinery in Light of the UAE Civil Code Compared to the Egyptian Civil Code* (Al-Halabi Legal Publications 2005) [in Arabic]

49. Vaccaro M and others, 'Advancing AI Negotiations: New Theory and Evidence from a Large-Scale Autonomous Negotiations Competition' (*arXiv*, 7 July 2025) arXiv:2503.06416v2. doi:10.48550/arXiv.2503.06416

50. Wachter S, Mittelstadt B and Floridi L, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) 7(2) International Data Privacy Law 76. doi:10.1093/idpl/ipx005

51. Zahra MAM, *Non-Administrative Sources of Obligation in the Civil Transactions Law of the United Arab Emirates: Harmful Acts and Beneficial Acts* (UAE University Press 2003) [in Arabic]

52. Zein S, 'The Civil Liability for Artificial Intelligence' (2023) 2022(1) BAU Journal of Legal Studies 14. doi:10.54729/2958-4884.1110

AUTHORS INFORMATION

Bashar Talal Momani*
PhD (Law), Assoc. Prof., Civil Law, University of Khorfakkan, Sharjah, United Arab Emirates
bashar.momani@ukf.ac.ae
https://orcid.org/0000-0002-0451-2551
Corresponding author, responsible for data curation, methodology, project administration, resources, supervision, validation, writing – original draft, writing – review & editing.

Nasr Farid Hassan
PhD (Law), Assoc. Prof., College of Law, Department of Private Law, Ajman University, Ajman, United Arab Emirates
n.farid@ajman.ac.ae
https://orcid.org/0000-0003-2648-9898
Co-author, responsible for data curation, methodology, resources, supervision, validation, writing – original draft, writing – review & editing.

Hosni Mahmoud AbdelDaiem AbdelSamad
PhD (Law), Assoc. Prof., University of Alazhar, Alazhar, Egypt
prof.hossny@outlook.com
https://orcid.org/0009-0000-5887-3195
Co-author, responsible for data curation, methodology, resources, supervision, validation, writing – original draft, writing – review & editing.

Mohamed Elsayed Eldessouky
PhD (Law), Assoc. Prof., College of Law, Dar Al Uloom University, Riyadh, Kingdom of Saudi Arabia
m.eldessouky@dau.edu.sa
https://orcid.org/0000-0001-9936-8996
Co-author, responsible for data curation, methodology, resources, supervision, validation, writing – original draft, writing – review & editing.

Competing interests: No competing interests were disclosed. Any potential conflict of interest must be disclosed by authors.

Disclaimer: The authors declare that their opinion and views expressed in this manuscript are free of any impact of any organizations.

RIGHTS AND PERMISSIONS

EDITORS

Managing Editor – Mag. Yuliia Hartman. **English Editor** – Julie Bold.
Ukrainian language Editor – Mag. Liliia Hartman.

ABOUT THIS ARTICLE

Cite this article
Momani BT, Hassan NF, AbdelDaiem AbdelSamad HM and Eldessouky ME, 'Legal Challenges Related to Contractual Negotiations via AI Technologies: Comparative Analytical Study' (2025) 8(Spec) Access to Justice in Eastern Europe 259-87 <https://doi.org/10.33327/AJEE-18-8.S-r000151>

DOI: https://doi.org/10.33327/AJEE-18-8.S-r000151

Keywords: *legal challenges, contractual negotiations, artificial intelligence (AI), civil liability, tort liability.*

DETAILS FOR PUBLICATION

Date of submission: 28 Apr 2025
Date of acceptance: 02 Sep 2025
Online First publication: 27 Oct 2025
Last Published: 30 Dec 2025
Whether the manuscript was fast tracked? - No
Number of reviewer report submitted in first round: 3 reports
Number of revision rounds: 2 rounds with minor revisions

Technical tools were used in the editorial process:
Plagiarism checks - Turnitin from iThenticate https://www.turnitin.com/products/ithenticate/
Scholastica for Peer Review https://scholasticahq.com/law-reviews

AI DISCLOSURE STATEMENT

The manuscript was prepared by the author. AI tools were employed exclusively for spelling, grammar, and stylistic refinement. No generative AI was used to produce original content, research ideas, or analysis.

АНОТАЦІЯ УКРАЇНСЬКОЮ МОВОЮ

Оглядова стаття

ПРАВОВІ ВИКЛИКИ, ПОВ'ЯЗАНІ З ПЕРЕГОВОРАМИ ЩОДО УКЛАДЕННЯ ДОГОВОРІВ, ПРОВЕДЕНИМИ ЗА ДОПОМОГОЮ ТЕХНОЛОГІЙ ШТУЧНОГО ІНТЕЛЕКТУ: ПОРІВНЯЛЬНО-АНАЛІТИЧНЕ ДОСЛІДЖЕННЯ

Башар Талал Момані*, Наср Фарід, Хусні Махмуд АбдельДаєм АбдельСамад та Мохамед Ельсаєд Ельдессукі

АНОТАЦІЯ

Вступ. *Переговори щодо укладення договорів, що проводяться системами штучного інтелекту (ШІ), спричиняють серйозні правові проблеми, зокрема порушують питання розподілу цивільної відповідальності за збитки, завдані через їхні помилки. Це дослідження, що використовує методологію порівняльного аналізу, виявляє значну прогалину в нормативному регулюванні в арабських юрисдикціях, зокрема в Єгипті та Об'єднаних Арабських Еміратах, де в законодавстві відсутні чіткі положення, що регулюють таку відповідальність. Натомість, нещодавні ініціативи Європейського Союзу, зокрема ризик-орієнтований підхід Закону про ШІ та нова система Директиви про відповідальність за ШІ, наголошують на відповідальності розробників та операторів.*

На цьому тлі в статті висловлюється підтримка розробки спеціалізованої арабської правової бази, яка б орієнтувалась на подібні моделі, зважаючи при цьому на місцеву специфіку. Така структура повинна містити: точне юридичне визначення інтелектуальних систем, конкретні механізми доказування для визначення вини та розподілу відповідальності, створення спеціального наглядового органу та зміцнення механізмів страхування як додаткових гарантій.

Методи. *Це дослідження використовує порівняльно-аналітичний метод для вивчення цивільної відповідальності за помилки ШІ в договірних переговорах, зосереджуючись на теоріях делікту та договорів згідно з єгипетським та еміратським законодавством, та порівнюючи їх з останніми розробками ЄС, зокрема Директивою про відповідальність за ШІ та Законом про дані, які надають чіткіші вказівки, ніж Закон про ШІ.*

Результати та висновки. *Порівняльний аналіз дає три основні результати. По-перше, існує явна прогалина в регуляторному полі в арабських юрисдикціях, які продовжують покладатися на загальні положення цивільного права без спеціалізованих правил для ШІ. По-друге, доктринальні відмінності між суворою відповідальністю в ЄС та ширшим підходом до відшкодування збитків в арабських системах ускладнюють будь-яке пряме перенесення європейських моделей. По-третє, проблеми з доказуванням залишаються центральними в обох системах, оскільки встановлення вини та причинно-наслідкового зв'язку у шкоді, пов'язаній зі ШІ, є за своєю суттю складним.*

Ключові слова: *правові виклики, договірні переговори, штучний інтелект (ШІ), цивільна відповідальність, деліктна відповідальність.*

ABSTRACT IN ARABIC

مقالة مراجعة

التحديات القانونية المتعلقة بالمفاوضات التعاقدية عبر تقنيات الذكاء الاصطناعي: دراسة تحليلية مقارنة

بشار طلال مومني*، نصر فريد حسن، حسني محمود عبد الدايم عبد الصمد، محمد السيد الدسوقي

الملخص

الخلفية: تُثير المفاوضات التعاقدية التي تُجرى بواسطة أنظمة الذكاء الاصطناعي (AI) تحديات قانونية عميقة، أبرزها مسألة تحديد المسؤولية المدنية عن الأضرار الناتجة عن أخطائها. تكشف هذه الدراسة، التي تعتمد المنهج التحليلي المقارن، عن فجوة تنظيمية واضحة في التشريعات العربية—وخاصة في مصر ودولة الإمارات العربية المتحدة—إذ تفتقر القوانين إلى نصوص صريحة تُنظّم مثل هذه المسؤولية. وفي المقابل، تُبرز المبادرات الأوروبية الحديثة، بما في ذلك النهج القائم على تقييم المخاطر في قانون الذكاء الاصطناعي (AI Act) والإطار الناشئ في توجيه مسؤولية الذكاء الاصطناعي (AI Liability Directive)، تركيزًا أساسيًا على مساءلة المطورين والمشغلين.

وفي هذا السياق، تدعو الورقة إلى تطوير إطار قانوني عربي متخصص يستلهم النماذج المقارنة مع الحفاظ على الخصوصيات المحلية. ينبغي أن يتضمن هذا الإطار ما يلي: تعريفًا قانونيًا دقيقًا للأنظمة الذكية، وآليات إثبات ملموسة لإسناد الخطأ وتوزيع المسؤولية، وإنشاء هيئة إشرافية متخصصة تُعنى بتنظيم ومراقبة استخدام تقنيات الذكاء الاصطناعي، إلى جانب تعزيز آليات التأمين بوصفها ضمانات تكميلية للحماية من المخاطر المحتملة.

المنهجية: تعتمد هذه الدراسة المنهج التحليلي المقارن في بحث المسؤولية المدنية عن أخطاء الذكاء الاصطناعي في المفاوضات التعاقدية، مع التركيز على نظريتي المسؤولية التقصيرية والعقدية في القانونين المصري والإماراتي، ومقارنتها بالتطورات الحديثة في الاتحاد الأوروبي، ولا سيّما توجيه مسؤولية الذكاء الاصطناعي (AI Liability Directive) وقانون البيانات (Data Act)، اللذين يقدّمان إرشادات أوضح وأكثر تحديدًا من قانون الذكاء الاصطناعي (AI Act).

النتائج والاستنتاجات: أفضى التحليل المقارن إلى ثلاث نتائج رئيسية. أولاً، هناك فجوة تنظيمية واضحة في الأنظمة القانونية العربية، إذ ما تزال تعتمد على الأحكام العامة للقانون المدني دون وجود قواعد متخصصة تنظم قضايا الذكاء الاصطناعي. ثانياً، إن الاختلافات المبدئية في الأسس القانونية بين مبدأ المسؤولية الصارمة المعتمد في الاتحاد الأوروبي، والمنهج العلاجي الأوسع المتبع في الأنظمة العربية، تجعل من نقل النماذج الأوروبية مباشرةً إلى السياق العربي أمرًا معقدًا وصعب التطبيق. ثالثاً، ما تزال التحديات المتعلقة بالإثبات تمثل محوراً جوهرياً في كلا النظامين، إذ إن إثبات الخطأ وعلاقة السببية في الأضرار المرتبطة بالذكاء الاصطناعي يُعدّ مسألة معقدة بطبيعتها نتيجة لطبيعة هذه الأنظمة واعتمادها على عمليات تحليلية مستقلة يصعب تتبعها بدقة.

Access to Justice in Eastern Europe
ISSN 2663-0575 (Print)
ISSN 2663-0583 (Online)
http://ajee-journal.com

Review Article

PRIVACY-BY-DESIGN IN EMOTION AI: DATA PROTECTION FRAMEWORKS AND COMPLIANCE STRATEGIES

Laroussi Chemlali* and Leila Benseddik

DOI:
https://doi.org/10.33327/AJEE-18-8.S-r000153

Date of submission: 21 Aug 2025
Date of acceptance: 02 Oct 2025
Online First publication: 04 Dec 2025
Last Published: 30 Dec 2025

ABSTRACT

Background: *The rapid advancement of Emotion Artificial Intelligence (Emotion AI) has created significant opportunities for innovation across a broad variety of domains, including healthcare, marketing, and human-computer interaction. Emotion AI applications—which process, analyse, and respond to human emotions—rely heavily on sensitive personal data, resulting in privacy and ethical concerns. The implementation of Privacy-by-Design (PbD) principles within such systems is essential to counter these challenges and maintain compliance with changing legal frameworks. This paper discusses the interplay between PbD and Emotion AI, with a special emphasis on the privacy risks associated with the collection and processing of emotional data. The study is set against the broader background of developing ethical AI, emphasising the urgent need to balance technological innovation with robust privacy protection.*

Methods: *The paper provides a conceptual legal analysis of the intersection between Privacy-by-Design (PbD) and Emotion AI within modern data protection frameworks. It employs a comprehensive review of primary sources, including the EU GDPR,*

the EU AI Act, CJEU and ECtHR jurisprudence, and guidance from Data Protection Authorities, alongside secondary sources like scholarly works and books. The discussion is structured first to provide an overview of Emotion AI, its applications, as well as individual privacy concerns it raises. This is followed by a consideration of existing data protection regimes and how they can be transferred to Emotion AI systems. The study then focuses on the fundamental principles of PbD, examining how they can be applied when developing and deploying Emotion AI technologies.

Results and Conclusions: *The analysis demonstrates that implementing PbD principles in Emotion AI systems is essential—not merely beneficial—for protecting users' privacy and ensuring legal compliance. Properly implemented PbD frameworks deliver three essential benefits: enhanced system transparency, stronger accountability mechanisms, and greater user control over their own data. These findings contribute significantly to the theoretical foundations of responsible AI design while offering actionable implementation guidance for organisations deploying Emotion AI systems. Finally, the study presents an unambiguous model for developers and organisations to successfully ride the wave of convergence between emotional intelligence technology and privacy regulations.*

1 INTRODUCTION

Emotion artificial intelligence technology, which can recognise and react to human emotions using information such as voice, facial expressions, and physiological signs, is a prevailing yet controversial invention as AI developments continue to transform sectors.[1] Although Emotion AI presents revolutionary prospects in healthcare, customer service, education, and other fields, it also poses serious questions around privacy, autonomy, and the ethical handling of sensitive emotional data. Traditional data protection paradigms are challenged by the capacity to collect and analyse such highly sensitive data, necessitating a strong and proactive reaction from companies, researchers, and regulators.[2]

This paper examines Privacy-by-Design (PbD) as a fundamental approach to addressing these ethical and legal issues. Stakeholders can guarantee adherence to data protection frameworks, including the General Data Protection Regulation (EU GDPR)[3] of the EU and comparable international regulations, as well as gain public trust in these cutting-edge

1 Andrew McStay, *Emotional AI: The Rise of Empathic Media* (SAGE Publications 2018) doi:10.4135/9781526451293.

2 Darlene Barker and others, 'Ethical Considerations in Emotion Recognition Research' (2025) 7(2) Psychology International 43. doi:10.3390/psycholint7020043; Andrew McStay, 'Emotional AI, Soft Biometrics and the Surveillance of Emotional Life: An Unusual Consensus on Privacy' (2020) 7(1) Big Data and Society 1. doi:10.1177/2053951720904386.

3 Regulation (EU) 2016/679 of the European Parliament and of the Council 'On the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation)' (27 April 2016) [2016] OJ L119/1.

technologies by incorporating privacy considerations into the development lifecycle of Emotion AI systems. The paper explores methods for implementing PbD principles into practice, while foreseeing future difficulties in striking a balance between human rights protection and technological progress.

By doing so, this paper aims to answer the following question: How can PbD principles be embedded into the development lifecycle of Emotion AI systems to foster robust data protection and compliance with privacy regulations? Therefore, this study offers a comprehensive understanding of how Emotion AI development can be aligned with contemporary compliance standards and ethical practices, ensuring a co-existence between innovation and privacy.

2 METHODOLOGY

This study was conducted as a conceptual legal analysis, aimed at understanding the intersection between the principle of PbD and the evolving use of Emotion AI within contemporary data protection frameworks. Although this study does not rely on empirical data collection, it employed a comprehensive review of primary and secondary legal materials. Primary sources included instruments, mainly the EU GDPR, along with the EU AI Act, ECtHR, and CJEU jurisprudence, as well as the Data Protection Authorities' Guidance. Secondary sources comprised mainly scholarly articles, books, and other relevant publications on the subject. The analysis proceeded in three stages. The paper presents a clarification of the conceptual foundations of PbD, drawing on its seven principles as articulated by Cavoukian. It then explores the implementation of the PbD in the Emotion AI setting. Finally, the paper was then followed by an examination of the compliance strategies proposed in existing frameworks, identifying practical enforcement challenges and gaps in regulatory guidance. By synthesising across legal systems and policy discourses, this conceptual analysis aimed not only to illustrate how PbD has been operationalised in the context of Emotion AI but also to point out where enforcement mechanisms fall short in practice.

3 UNDERSTANDING EMOTION AI AND ITS PRIVACY AND ETHICAL CONCERNS

3.1. Defining Emotion AI

Combining artificial intelligence with the intricate realm of human emotions, Emotion AI forms an intriguing link between technology and psychology. This field has increasingly evolved since Rosalind Picard's groundbreaking research in 1995,[4] and currently includes a variety of technologies that are able to recognise, understand, and react to human emotional

4 Rosalind W Picard, *Affective Computing* (Technical Report no 321, MIT Media Laboratory Perceptual Computing Section 1995).

states. To develop systems that can successfully interpret human emotional states, this complex field integrates machine learning, computer vision, and psychological insights.[5]

These systems analyse several signs at once, including subtle bodily movements, variations in voice patterns, minor changes in facial expressions, and quantifiable physiological reactions, such as alterations in skin conductance and heart rhythm fluctuations.[6] Through this inclusive approach, abstract emotional experiences can be converted into concrete, analysable data points that computers can process systematically.[7]

Emotion AI's real-world uses have grown significantly in a variety of industries. Measuring consumer responses to goods and services allows for more individualised shopping experiences in retail.[8] These technologies are used by medical professionals to track patients' emotional states throughout therapy and to spot early signs of mental health issues.[9] Emotion AI also plays a role in security applications, analysing crowd behaviour and detecting threats.[10]

The pervasive integration of Emotion AI highlights its profound impact on virtually every facet of human life. In fact, virtual assistants can now identify emotional undertones in speech patterns and modify their responses accordingly.[11] To improve comfort and safety, modern cars can now be equipped with technologies that monitor drivers' emotional states.[12] Smartphones increasingly feature advanced emotion-detection capabilities through their cameras and microphones.[13] Additionally, there has been a notable acceleration in the institutional adoption of Emotion AI: educational institutions use it to monitor students'

5 Sesha Bhargavi Velagaleti and others, 'Empathetic Algorithms: The Role of AI in Understanding and Enhancing Human Emotional Intelligence' (2024) 20(3) Journal of Electrical Systems 2051. doi:10.52783/jes.1806.

6 Smith K Khare and others, 'Emotion Recognition and Artificial Intelligence: A Systematic Review (2014-2023) and Research Recommendations' (2024) 102 Information Fusion 102019. doi:10.1016/j.inffus.2023.102019.

7 Bei Pan and others, 'A Review of Multimodal Emotion Recognition from Datasets, Preprocessing, Features, and Fusion Methods' (2023) 561 Neurocomputing 126866. doi:10.1016/j.neucom.2023.126866.

8 Thomas Davenport and others, 'How Artificial Intelligence Will Change the Future of Marketing' (2020) 48(1) Journal of the Academy of Marketing Science 24. doi:10.1007/s11747-019-00696-0.

9 Anoushka Thakkar, Ankita Gupta and Avinash De Sousa, 'Artificial Intelligence in Positive Mental Health: A Narrative Review' (2024) 6 Frontiers in Digital Health 1280235. doi:10.3389/fdgth.2024.1280235.

10 Lena Podoletz, 'We Have to Talk A bout Emotional AI and Crime' (2023) 38 AI & Society 1067. doi:10.1007/s00146-022-01435-w.

11 Ruhul Amin Khalil and others, 'Speech Emotion Recognition Using Deep Learning Techniques: A Review' (2019) 7 IEEE Access 117327. doi:10.1109/ACCESS.2019.2936124.

12 Sebastian Zepf and others, 'Driver Emotion Recognition for Intelligent Vehicles: A Survey' (2020) 53(3) ACM Computing Surveys (CSUR) 1. doi:10.1145/338879.

13 Imran A Zualkernan and others, 'Emotion Recognition Using Mobile Phones' (2017) 60 Computers & Electrical Engineering 1. doi:10.1016/j.compeleceng.2017.05.004.

emotional health and level of engagement during class activities,[14] while healthcare facilities employ it to monitor the psychological moods and recovery status of their patients.[15] In the same vein, corporate environments utilise these technologies to gauge employee engagement and workplace satisfaction.[16]

The extensive use of Emotion AI across diverse fields highlights both its possible advantages and the necessity of carefully weighing its effects on individual autonomy and privacy.

3.2. The Rise of Privacy and Ethical Concerns with Emotion AI

Although developers emphasise the importance of anonymity and collective emotional analysis, the collection of emotional data still poses serious privacy issues. While there might exist other factors, this section examines the issues and the implications associated with the collection of emotional data.

3.2.1. Bias and Discrimination

One of the most critical issues in AI is the ethical implications of bias and discrimination in Emotion AI systems. The training data these systems rely on has a fundamental impact on their performance. When this data includes pre-existing societal biases or prejudices, the emerging AI systems unavoidably reflect—and may even reinforce—these discriminatory patterns.[17]

When they arise, such biases can significantly impact different social and demographic groups. Emotion AI systems, in particular, may unfairly impact groups based solely on how the AI interprets their emotional responses.[18] This could then result in difficult situations where the technological tools intended to improve human connection end up being used in a discriminatory way.

Beyond racial biases, this issue includes cultural variations in emotional expression. For instance, Emotion AI systems that are usually trained on data from dominant cultural

14 Angel Olider Rojas Vistorte and others, 'Integrating Artificial Intelligence to Assess Emotions in Learning Environments: A Systematic Literature Review' (2024) 15 Frontiers in Psychology 1387089. doi:10.3389/fpsyg.2024.1387089.

15 Prashant Kumar Nag, Amit Bhagat and R Vishnu Priya, 'Expanding AI's Role in Healthcare Applications: A Systematic Review of Emotional and Cognitive Analysis Techniques' [2025] IEEE Access. doi:10.1109/ACCESS.2025.3562131.

16 McStay (n 2).

17 Varsha PS, 'How Can We Manage Biases in Artificial Intelligence Systems: A Systematic Literature Review' (2023) 3(1) International Journal of Information Management Data Insights 100165. doi:10.1016/j.jjimei.2023.100165.

18 Nomisha Kurian, 'AI's Empathy Gap: The Risks of Conversational Artificial Intelligence for Young Children's Well-Being and Key Ethical Considerations for Early Childhood Education and Care' (2023) 26(1) Contemporary Issues in Early Childhood 132. doi:10.1177/14639491231206004.

groups may struggle to accurately identify and interpret emotional responses from other cultures. In other words, when Emotion AI is deployed across various cultural contexts, its underlying cultural bias may lead to misinterpretations and inaccurate assessments.[19]

3.2.2. Transparency and Explainability

To build trust and accountability in human-machine interactions, Emotion AI systems must be transparent and understandable. Indeed, trust is a necessary condition for the effective deployment and uptake of Emotion AI technologies rather than just a desirable attribute.[20] In the absence of a solid foundation of trust, the most complex and powerful Emotion AI systems might fail to accomplish their purposes.

Therefore, maintaining and preserving trust requires Emotion AI systems to function with a high level of transparency in a number of crucial areas. This includes being transparent about how the system functions, how it processes information, and—above all—how it manages sensitive emotional data. Emotion AI users need a thorough awareness of the complete data lifecycle, from the initial phase of emotional reactions to the storage processes and the final uses of this data.[21]

However, recent studies demonstrate that many Emotion AI systems in use today function as "black boxes," with internal procedures that are frequently hidden from both users and observers.[22] This lack of transparency presents serious challenges. Consumers struggle to trust these systems' outputs or evaluate their reliability when they cannot comprehend how decisions are made. Furthermore, this opacity makes it more difficult to detect and resolve potential issues with the systems, such as algorithmic biases or systematic errors.[23]

This transparency issue has implications that go beyond user confidence. It affects the wider accountability of Emotion AI systems and raises significant questions about their responsible creation and application. Clear, intelligible, and transparent systems are

19 Peter Mantello and others, 'Machines that Feel: Behavioral Determinants of Attitude Towards Affect Recognition Technology—Upgrading Technology Acceptance Theory with the Mind sponge Model' (2023) 10(1) Humanities and Social Sciences Communications 430. doi:10.1057/s41599-023-01837-1.

20 Keng L Siau and Weiyu Wang, 'Building Trust in Artificial Intelligence, Machine Learning, and Robotics' (2018) 31(2) Cutter Business Technology Journal 47.

21 Ben Chester Cheong, 'Transparency and Accountability in AI Systems: Safeguarding Wellbeing in the Age of Algorithmic Decision-Making' (2024) 6 Frontiers in Human Dynamics 1421273. doi:10.3389/fhumd.2024.1421273.

22 Angelica Salvi del Pero, Peter Wyckoff and Ann Vourc'h, *Using Artificial Intelligence in the Workplace: What are the Main Ethical Risks?* (Social, Employment and Migration Working Papers no 273, OECD 2022). doi:10.1787/840a2d9f-en.

23 Karina Cortiñas-Lorenzo and Gerard Lacey, 'Toward Explainable Affective Computing: A Review' (2023) 35(10) IEEE Transactions on Neural Networks and Learning Systems 13101. doi:10.1109/TNNLS.2023.3270027.

becoming increasingly necessary as these technologies permeate more facets of our lives, ensuring they genuinely serve the interests of both their users and society as a whole.[24]

3.2.3. Consent and Data Security

A fundamental element of the ethical collection and use of personal information—especially emotional information—is informed consent. In the current digital environment, where Emotion AI systems are growing more common, this basic requirement has grown more complex. The alarming issue, however, is the lack of awareness among people regarding their emotions, which are being monitored and examined, let alone the possible consequences of such monitoring.[25] This lack of awareness is particularly concerning in situations where Emotion AI tools are used covertly or when people feel pressured to give their consent, such as a condition of employment or to access specific services.[26]

3.2.4. The Risk of Manipulation

There are serious ethical issues regarding Emotion AI's ability to assess and predict human emotions, particularly in relation to manipulation. The capacity of this technology to interpret physiological signs, speech patterns, facial expressions, and behavioural data opens up possibilities for impacting human behaviour.[27] Businesses, for instance, could employ Emotion AI in advertising to develop highly targeted advertisements that take advantage of people's emotional weaknesses.[28] An algorithm might, for example, detect when a person feels lonely or insecure and display advertisements intended to capitalise on these feelings. This could lead to impulsive purchasing decisions or even dependence on particular goods or services.[29]

24 Alejandro Barredo Arrieta and others, 'Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI' (2020) 58 Information Fusion 82. doi:10.1016/j.inffus.2019.12.012.

25 Adam J Andreotta, 'The Hard Problem of AI Rights' (2021) 36 AI & Society 19. doi:10.1007/s00146-020-00997-x

26 Adam J Andreotta, Nin Kirkham and Marco Rizzi, 'AI, Big Data, and the Future of Consent' (2022) 37 AI & Society 1715. doi:10.1007/s00146-021-01262-5.

27 Marcello Ienca, 'On Artificial Intelligence and Manipulation' (2023) 42(3) Topoi 833. doi:10.1007/s11245-023-09940-3.

28 V Kumar and others, 'Understanding the Role of Artificial Intelligence in Personalized Engagement Marketing' (2019) 61(4) California Management Review 135. doi:10.1177/0008125619859317.

29 Andrew McStay, 'Empathic Media and Advertising: Industry, Policy, Legal and Citizen Perspectives (The Case for Intimacy)' (2016) 3(2) Big Data & Society. doi:10.1177/2053951716666868.

4 DATA PROTECTION FRAMEWORKS RELEVANT TO EMOTION AI

Existing data protection standards, along with newly emerging AI-specific laws and regulations, have created a complex web of compliance requirements, making the legal environment governing Emotion AI more comprehensive. Considering that the EU has the world's most comprehensive and stringent data protection and AI standards, serving as a benchmark, this section focuses primarily on the EU GDPR and the EU Artificial Intelligence Act—both of which serve as key instruments governing data protection and Emotion AI.

4.1. Emotion AI Through the Lenses of the GDPR

The GDPR serves as the cornerstone legislation governing personal data processing within the European Union. Its jurisdiction transcends geographical boundaries, extending to any entity, regardless of location, that monitors the behaviour of individuals in the EU.

Article 9(1) of the GDPR expressly prohibits the "processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation". Although emotional data is not explicitly listed in Article 9(1), there is a broad consensus that it qualifies as biometric data when it meets the personal data criteria outlined in Article 4(1) of the GDPR.[30]

As a result, companies using Emotion AI systems must carefully adhere to the GDPR's core requirements of purpose limitation, data minimisation, and storage limitation. This means restricting data collection to the most critical components, providing a clear explanation for the acquisition of emotional data, and retaining such data only for as long as necessary. The regulation also requires enterprises to disclose their use of Emotion AI in a transparent and understandable manner, while ensuring that data subjects have accessible means to manage their personal data.

To ensure regulatory compliance, organisations must also put in place thorough measures that combine organisational and technical safeguards. These include modern encryption techniques, strictly regulated access controls, and thorough documentation of all data processing operations, which should be part of these safeguards. Additionally, companies must conduct comprehensive Data Protection Impact Assessments to identify and mitigate potential risks before implementing large-scale Emotion AI systems that handle biometric data. By guaranteeing that privacy considerations are incorporated into the system's design from the beginning, these assessments act as essential preventive measures. In addition to demonstrating a commitment to compliance, adopting stringent security measures helps protect private emotional information from unethical or illegal use.

30 Leonhard Menges and Eva Weber-Guskar, 'Digital Emotion Detection, Privacy, and the Law' (2025) 38(2) Philosophy & Technology 1. doi:10.1007/s13347-025-00895-4.

4.2. Emotion AI Systems Under the EU AI Act

In 2024, the EU put forward an AI regulation,[31] marking the initial steps towards regulating AI and transforming the legal assessment of AI systems. This regulation uses a risk-based approach, categorising AI systems into four levels—unacceptable, high, limited, and minimal risk—based on the risk they pose to people and society.[32] Depending on its classification, every category of AI system is subject to specific rules and restrictions.

This thoughtful approach reflects a sophisticated understanding of the diverse implications that different AI technologies may have. The level of regulatory monitoring is directly correlated with the possible harm associated with each AI application under the hierarchical oversight framework established by the AI Act. The regulation thus establishes distinct lines between permissible and impermissible AI systems.

According to Article 5 (1)(f) of the AI Act, "the placing on the market, the putting into service for this specific purpose, or the use of AI systems to infer emotions of a natural person in the areas of workplace and education institutions, except where the use of the AI system is intended to be put in place or into the market for medical or safety reasons" is strictly prohibited. This rationale for this prohibition is further detailed in Recital 44, which lists several serious concerns regarding emotion identification systems. These concerns include its intrusive nature, lack of specificity and generalisability, power imbalance between those using it and those affected, its limited reliability, and the potential for discriminatory outcomes.[33]

Furthermore, Annexe III of the AI Act classifies emotion recognition systems as high-risk AI systems, subjecting them to stringent regulatory requirements. As a result, those who implement such systems must adhere to Article 50(3) of the AI Act, which stipulates that individuals who are exposed to these technologies must be made aware of how they work and that any personal information they handle must be processed in full compliance with the requirements of the GDPR.[34]

4.3. European Case Law and Regulatory Guidance on Emotion AI

31 Regulation (EU) 2024/1689 of the European Parliament and of the Council 'Laying Down Harmonised Rules on Artificial Intelligence and Amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)' (13 June 2024) <https://eur-lex.europa.eu/eli/reg/2024/1689/oj> accessed 23 September 2025.

32 Natalia Díaz-Rodrígue and others, 'Connecting the Dots in Trustworthy Artificial Intelligence: From AI Principles, Ethics, and Key Requirements to Responsible AI Systems and Regulation' (2023) 99 Information Fusion 101896. doi:10.1016/j.inffus.2023.101896.

33 See: Regulation (EU) 2024/1689 (n 31) recital 44.

34 ibid, art 50(3).

The use of biometric and emotion-inference technologies has been severely limited by European jurisprudence and supervisory guidelines, which stress the principles of necessity, proportionality and periodic review. The Court of Justice of the European Union (CJEU) has frequently ruled that national schemes authorising the systematic or blanket collection of biometric and genetic data violate the EU data-protection frameworks unless strictly limited and justified.

This principle was clearly articulated in the V.S. case, where the CJEU held that compulsory, systematic collection of biometric data for police records must meet a "strictly necessary" requirement in pursuit of specific, legitimate objectives.[35] Another landmark decision, *Digital Rights Ireland Ltd v Minister,* declared the EU Data Retention Directive invalid on the grounds that mass and indiscriminate retention of communications metadata without adequate protection is a severe interference with rights to data privacy.[36] According to this case, any automated or intrusive processing of personal data, especially sensitive or behavioural data, must adhere to stringent requirements for necessity, proportionality, transparency, and legal justification. Emotion AI, insofar as it infers inner emotional states, implicates several of those concerns.

In parallel, the ECtHR also considered intrusive biometric surveillance as a matter falling under Article 8 (private life). In *Glukhin v. Russia*, the ECtHR ruled that the use of facial-recognition technology by authorities was seen as "highly intrusive", highlighting that the use poses serious proportionality and clarity issues in domestic law.[37] Earlier, Gaughran v. the United Kingdom highlighted the dangers of stigmatisation and life-course harm and criticised schemes that permitted the indefinite retention of biometric data (photographs and fingerprints) as being incompatible with Convention provisions.[38]

Supervisory authorities have converted these principles into operational guidance. The European Data Protection Board's guidelines on facial recognition and law enforcement processing require stringent purpose limitation, data minimisation, and human oversight safeguards before any biometric processing. They also call for customised necessity

35 Case C-205/21 *VS v Ministerstvo na vatreshnite raboti, Glavna direktsia za borba s organiziranata prestapnost* (CJEU, 26 January 2023) <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A62021CJ0205> accessed 23 September 2025.

36 Joined Cases C-293/12 and C-594/12 *Digital Rights Ireland Ltd v Minister for Communications, Marine and Natural Resources & Others* (**CJEU** (Grand Chamber), 8 April 2014) <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A62012CA0293> accessed 23 September 2025.

37 *Glukhin v Russia* App no 11519/20 (ECtHR, 4 July 2023) **<**https://hudoc.echr.coe.int/eng?i=001-225655**>** accessed 23 September 2025**. See:** Francesca Palmiotto and Natalia Menéndez González, 'Facial Recognition Technology, Democracy and Human Rights' (2023) 50 *Computer Law & Security Review* 105857. doi:10.1016/j.clsr.2023.105857.

38 *Gaughran v the United Kingdom* App no 45245/15 (ECtHR, 13 February 2020) **<**https://hudoc.echr.coe.int/eng?i=001-200817**>** accessed 23 September 2025.

assessments rather than general authorisations.[39] Similarly, the European Data Protection Supervisor identified face emotion recognition as especially risky, pointing out the strong potential for discriminatory consequences and scientific ambiguity over accuracy.[40]

At the National DPA level, the UK Information Commissioner's Office (ICO) has frequently cautioned about the immaturity, bias, and intrusiveness of emotion-recognition technology. These concerns have been incorporated into the ICO's AI strategy and biometrics guidance.[41] In a similar vein, France's CNIL[42] and Spain's AEPD[43] published white papers and technical dispatches addressing automatic processing (speech, face, and neurodata) and underlining the concerns where inferred mental states meet with "special categories" of data. Regulators emphasise DPIAs, necessity, transparency, and purpose limitation. They also frequently caution against covert or workplace applications that lack robust protections.

5 COMPLIANCE STRATEGIES FOR PRIVACY-BY-DESIGN IN EMOTION AI

PbD emphasises the importance of incorporating privacy considerations throughout the development lifecycle to reduce risks and boost user confidence.[44] The present section offers a comprehensive insight into the underlying core principles of PbD and explores their implementation approaches in Emotion AI.

39 European Data Protection Board, *Guidelines 05/2022 on the Use of Facial Recognition Technology in the Area of Law Enforcement* (version 2.0, 26 April 2023) <https://www.edpb.europa.eu/system/files/2023-05/edpb_guidelines_202304_frtlawenforcement_v2_en.pdf?utm_source=chatgpt.com> accessed 23 September 2025.

40 '*TechDispatch #1/2024 – Neurodata*' (*European Data Protection Supervisor*, 3 June 2024) <https://www.edps.europa.eu/data-protection/our-work/publications/techdispatch/2024-06-03-techdispatch-12024-neurodata>accessed 26 September 2025.

41 '*Regulating AI: The ICO's Strategic Approach*' (*Information Commissioner's Office*, 30 April 2024) <https://ico.org.uk/about-the-ico/consultations/regulating-ai-the-icos-strategic-approach-a-response-to-the-dsit-secretary-of-state/> accessed 23 September 2025.

42 'Artificial Intelligence: the Opinion of the CNIL and its Counterparts on the Future European Regulation' (*CNIL Commission Nationale de l'Informatique et des Libertés*, 18 June 2021) <https://www.cnil.fr/en/artificial-intelligence-opinion-cnil-and-its-counterparts-future-european-regulation> accessed 23 September 2025.

43 TechDispatch (n 40).

44 Ann Cavoukian, 'Privacy by Design: The 7 Foundational Principles' (2009) 5 Information and Privacy Commissioner of Ontario, Canada 12.

5.1. Privacy-by-Design Core Principles

PbD is a forward-thinking framework for safeguarding personal data, embedding privacy considerations into the design specifications of technologies, systems, and processes at the outset of personal data processing, rather than as an afterthought. This concept is further clarified by Cavoukian, who outlines seven fundamental principles that serve as a valuable reference for organisations seeking to comply with data protection regulations.[45] These are:

- **Proactive, not Reactive**: This approach anticipates and prevents privacy issues before they occur, rather than addressing them after violations have happened. It helps organisations maintain trust and avoid costly remediation efforts.
- **Privacy as the Default Setting**: Systems should safeguard user privacy automatically, without requiring user intervention. Every business procedure or IT system should automatically, by default, protect personal data.
- **Privacy Embedded into Design**: The system's architecture and design incorporate privacy protection from the start, rather than adding it after. This guarantees that privacy becomes a core feature.
- **Full Functionality – Positive-Sum, not Zero-Sum**: PbD seeks to accommodate all legitimate interests and objectives in a mutually beneficial "win-win" manner, rather than relying on an outdated zero-sum approach that requires needless trade-offs.
- **End-to-end Security**: Stringent procedures must be applied throughout the entire data lifecycle. This guarantees that all the data is securely collected, used, retained, and deleted at the end of the process.
- **Visibility and Transparency**: All components and operations should remain visible and transparent to both users and providers, fostering trust and accountability.
- **Respect for User Privacy**: Individuals' interests are crucial, requiring robust privacy defaults, and user-friendly options, as well as proper notice.

In recent years, the concept of PbD has gained widespread acceptance and has been formally incorporated into major data protection frameworks such as the GDPR. The principle is emphasised in Recital 78 of the GDPR, which states that:

> "When developing, designing, selecting, and using applications, services, and products that involve personal data processing, developers and manufacturers should consider the right to data protection, adhering to the state of the art, and ensuring that controllers and processors fulfil their obligations."[46]

45 ibid.

46 See: Regulation (EU) 2016/679 (n 3) recital 78.

Article 25(1) of the GDPR provides further guidance on implementing PbD. It requires controllers to take organisational and technical steps to implement data protection principles, such as data minimisation, and to incorporate the necessary safeguards to meet regulatory requirements and uphold individuals' rights. These measures should take into account various factors, including technological advancements, associated costs, and the nature, scope, and potential risks involved in data processing activities.

5.2. Implementing Privacy-by-Design in Emotion AI

PbD implementation in Emotion AI requires a thorough policy that covers every component of the technology, from data collection and processing to storage and utilisation.

5.2.1. Data Minimisation and Purpose Limitation

Fundamental PbD principles of purpose limitation and data minimisation necessitate that companies collect just the data necessary for a specific, legitimate purpose and use it only for that purpose. Therefore, applying PbD principles to emotional data poses substantial real-world challenges from an enforcement standpoint. Organisations must operationalise data minimisation and purpose limitation across intricate technical systems and business processes while navigating data protection regulations.

In practice, organisations must keep thorough records demonstrating their assessment of the importance of each emotional data point gathered. For instance, a mental health app would need to explain why specific emotional indicators are necessary for delivering therapeutic services, while omitting information that might be relevant but not essential for the service. This documentation becomes essential when Data Protection Authorities conduct regulatory audits and investigations.

The challenge of purpose limitation is reflected in data access controls and system architecture. Organisations must implement technical safeguards to prevent the use of emotional data for secondary purposes, such as product development or marketing, unless this is stated explicitly in the original consent. Addressing this challenge often requires sophisticated data governance frameworks and regular compliance audits. For instance, a workplace wellness program collecting emotional data must ensure strict separation between health-related processing and performance management systems to maintain compliance and protect individuals' privacy.

5.2.2. Transparency and Consent

Implementing informed consent for emotional data within EU frameworks poses several practical challenges. Organisations should set up clear, accessible consent mechanisms that effectively clarify the intricacies of emotional AI systems, while ensuring users are not overwhelmed by excessive consent requests. Information on an organisation's data collection

and processing procedures, including the type of emotional data it collects, its intended uses, and the security measures in place, should be easily comprehensible and available.[47]

The technical implementation of data subject rights raises significant challenges, particularly for emotional data that may be integrated into multiple systems or used for training AI models. For organisations to track emotional data throughout their infrastructure, fulfil access requests, and implement the right to be forgotten, they require strong data mapping and management systems. This becomes especially complex when emotional data is derived from multiple sources or combined with other personal data.

As emotional AI technologies evolve, organisations need to maintain ongoing transparency. For this, they are required to regularly update their documentation and communication channels. This includes incorporating practical processes for consent withdrawal at any time and the ability to view and request changes to their data.[48]

5.2.3. Security and Data Protection

Adopting thorough data security procedures is another essential component for protecting sensitive information. These precautions should ensure that this private information is protected from exposure, misuse, and unauthorised access.[49] Indeed, ways to implement such precautions include using advanced encryption methods, establishing strict access controls, and conducting systematic audits of data protection frameworks to identify weaknesses.[50] By prioritising these initiatives, organisations may improve accountability, lower risks, and maintain ethical principles while managing Emotion AI systems as well as protecting people's rights and privacy.

5.2.4. Accountability and Algorithmic Transparency

To ensure that Emotion AI systems function in a way that avoids discrimination, maintain fairness and stop bias, algorithmic accountability and transparency are crucial.[51] Thus, to make these algorithms as transparent as possible, organisations should provide a thorough, understandable explanation of the procedures and methods that underpin their operation.

47 Yugang Li and others, 'Developing Trustworthy Artificial Intelligence: Insights from Research on Interpersonal, Human-Automation, and Human-AI Trust' (2024) 15 Frontiers in Psychology 1382693. doi:10.3389/fpsyg.2024.1382693; Barker and others (n 2).

48 Salvi del Pero, Wyckoff and Vourc'h (n 22).

49 Ramanpreet Kaur, Dušan Gabrijelčič and Tomaž Klobučar, 'Artificial Intelligence for Cybersecurity: Literature Review and Future Research Directions' (2023) 97 Information Fusion 101804. doi:10.1016/j.inffus.2023.101804.

50 Jon D Elhai, Jason C Levine and Brian J Hall, 'Anxiety about Electronic Data Hacking: Predictors and Relations with Digital Privacy Protection Behavior' (2017) 27(3) Internet Research 631. doi:10.1108/IntR-03-2016-0070.

51 Julianne Cartwright, Marcus T Ellington and Amelia S Hawthorne, 'Exploring the Potential of Emotional Intelligence in AI through Emotion-Sensitive Learning Algorithms' [2024] Preprint. doi:10.13140/RG.2.2.18780.16005.

Transparency challenges derive from the complexity of AI systems, especially with Emotion AI's subjective nature. Organisations must address this by providing clear documentation and illustrative insights about how algorithms avoid bias, handle emotional input, and justify outputs, without overloading stakeholders with technical jargon.

Promoting accountability in Emotion AI systems requires not only transparency but also clarity on decision-making procedures. Companies should specify exactly how particular decisions are made using the algorithm, including the inputs and the justification for the results. By doing this, companies promote confidence among stakeholders such as regulators, users, and the public, in addition to enabling improved oversight.[52] Therefore, implementing such procedures strengthens the integrity and social responsibility of the Emotion AI systems.

5.2.5. User-Centric Design

User-centric design prioritises individuals' needs, preferences, and rights in the development and deployment of Emotion AI systems.[53] This approach necessitates continuous user engagement throughout the design processes to ensure that technologies align with human-centric values and respect user experiences. Practical enforcement involves integrating user feedback loops into the system development process. This could be done by facilitating participatory workshops or usability testing to enhance system responsiveness and meet user expectations. Designers must balance innovation with compliance, ensuring that emotional data is handled ethically, securely, and lawfully.

Furthermore, user-centric design empowers individuals with greater control over their emotional data. This includes enabling users to manage their data in ways that have personal significance, providing access to their emotional data, and offering simple methods to correct errors or delete the data if they choose. By prioritising consumers' autonomy and privacy first, this approach promotes both fairness and increases users' trust in the technology.

5.2.6. Impact Assessments

For businesses using Emotion AI systems, Data Protection Impact Assessments (DPIAs) are essential systematic methods for assessing and mitigating potential privacy issues prior to their deployment.[54] These assessments include a thorough examination of the

52 Cheong (n 21).

53 Lakshita Dodeja and others, 'Towards the Design of User-Centric Strategy Recommendation Systems for Collaborative Human–AI Tasks' (2024) 184 International Journal of Human-Computer Studies 103216. doi:10.1016/j.ijhcs.2023.103216.

54 Denise Almeida, Konstantin Shmarko and Elizabeth Lomas, 'The Ethics of Facial Recognition Technologies, Surveillance, and Accountability in an Age of Artificial Intelligence: A Comparative Analysis of US, EU, and UK Regulatory Frameworks' (2022) 2 AI Ethics 377. doi:10.1007/s43681-021-00077-w; Margot E Kaminski and Gianclaudio Malgieri, 'Algorithmic Impact Assessments Under the GDPR: Producing Multi-Layered Explanations' (2021) 11(2) International Data Privacy Law 125. doi:10.1093/idpl/ipaa020.

methods used to collect, process, store, and use personal and emotional data while taking into account the particular sensitivities related to Emotion AI technology. A comprehensive DPIA typically considers various factors, including the necessity and appropriateness of data processing, potential threats to people's rights and freedoms, and the effectiveness of existing protections.

DPIAs can be used by organisations to identify certain vulnerabilities, including dangers of emotional manipulation, potential bias or discrimination in emotional analysis, and illegal access to emotional data. Based on these results, organisations can create focused mitigation strategies, like enhanced data encryption, stringent access controls, explicit data retention guidelines, and transparent and honest communication with users regarding the processing of their emotional data. DPIAs are a dynamic tool for upholding ethical norms and privacy compliance, and regular revisions are crucial as Emotion AI technology develops and new privacy risks appear.

6 CONCLUSIONS

Incorporating PbD principles into the development of Emotion AI is not only required by law but also essential to the long-term, ethical development of this technology. As this paper demonstrates, the sensitive nature of emotional data necessitates a thorough, proactive approach to privacy protection that goes beyond mere compliance with current regulations. Implementing PbD in Emotion AI systems requires careful consideration of numerous factors, including data minimisation, transparency, consent, and robust security mechanisms. Therefore, organisations must strike a balance between the innovation of Emotion AI and the fundamental right to privacy to prevent technological innovation from compromising individual autonomy and trust.

The development and implementation of Emotion AI systems will likely be influenced by future regulatory changes, especially with the implementation of the EU AI Act and the ongoing impact of the GDPR. Businesses will be better positioned to adapt to these shifting demands while upholding public trust if they adopt PbD principles early in their development process.

Finally, organisations can develop solutions that are not just secure and compliant but also ethical and with long-lasting effects by integrating privacy concerns into the foundation of emotion AI systems. To ensure that the advancement of Emotion AI aligns with social values and expectations while encouraging innovation in this rapidly evolving sector, it will be crucial to conduct ongoing research, collaborate with stakeholders, and regularly evaluate privacy standards.

REFERENCES

1. Almeida D, Shmarko K and Lomas E, 'The Ethics of Facial Recognition Technologies, Surveillance, and Accountability in an Age of Artificial Intelligence: A Comparative Analysis of US, EU, and UK Regulatory Frameworks' (2022) 2 AI Ethics 377. doi:10.1007/s43681-021-00077-w
2. Andreotta AJ, 'The Hard Problem of AI Rights' (2021) 36 AI & Society 19. doi:10.1007/s00146-020-00997-x
3. Andreotta AJ, Kirkham N and Rizzi M, 'AI, Big Data, and the Future of Consent' (2022) 37 AI & Society 1715. doi:10.1007/s00146-021-01262-5
4. Barker D and others, 'Ethical Considerations in Emotion Recognition Research' (2025) 7(2) Psychology International 43. doi:10.3390/psycholint7020043
5. Barredo Arrieta A and others, 'Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI' (2020) 58 Information Fusion 82. doi:10.1016/j.inffus.2019.12.012
6. Cartwright J, Ellington MT and Hawthorne AS, 'Exploring the Potential of Emotional Intelligence in AI through Emotion-Sensitive Learning Algorithms' [2024] Preprint. doi:10.13140/RG.2.2.18780.16005
7. Cavoukian A, 'Privacy by Design: The 7 Foundational Principles' (2009) 5 Information and Privacy Commissioner of Ontario, Canada 12
8. Cheong BC, 'Transparency and Accountability in AI Systems: Safeguarding Wellbeing in the Age of Algorithmic Decision-Making' (2024) 6 Frontiers in Human Dynamics 1421273. doi:10.3389/fhumd.2024.1421273
9. Cortiñas-Lorenzo K and Lacey G, 'Toward Explainable Affective Computing: A Review' (2023) 35(10) IEEE Transactions on Neural Networks and Learning Systems 13101. doi:10.1109/TNNLS.2023.3270027
10. Davenport T and others, 'How Artificial Intelligence Will Change the Future of Marketing' (2020) 48(1) Journal of the Academy of Marketing Science 24. doi:10.1007/s11747-019-00696-0
11. Díaz-Rodrígue N and others, 'Connecting the Dots in Trustworthy Artificial Intelligence: From AI Principles, Ethics, and Key Requirements to Responsible AI Systems and Regulation' (2023) 99 Information Fusion 101896. doi:10.1016/j.inffus.2023.101896
12. Dodeja L and others, 'Towards the Design of User-Centric Strategy Recommendation Systems for Collaborative Human–AI Tasks' (2024) 184 International Journal of Human-Computer Studies 103216. doi:10.1016/j.ijhcs.2023.103216

13. Elhai JD, Levine JC and Hall BJ, 'Anxiety about Electronic Data Hacking: Predictors and Relations with Digital Privacy Protection Behavior' (2017) 27(3) Internet Research 631. doi:10.1108/IntR-03-2016-0070

14. Ienca M, 'On Artificial Intelligence and Manipulation' (2023) 42(3) Topoi 833. doi:10.1007/s11245-023-09940-3

15. Kaminski ME and Malgieri G, 'Algorithmic Impact Assessments Under the GDPR: Producing Multi-Layered Explanations' (2021) 11(2) International Data Privacy Law 125. doi:10.1093/idpl/ipaa020

16. Kaur R, Gabrijelčič D and Klobučar T, 'Artificial Intelligence for Cybersecurity: Literature Review and Future Research Directions' (2023) 97 Information Fusion 101804. doi:10.1016/j.inffus.2023.101804

17. Khalil RA and others, 'Speech Emotion Recognition Using Deep Learning Techniques: A Review' (2019) 7 IEEE Access 117327. doi:10.1109/ACCESS.2019.2936124

18. Khare SK and others, 'Emotion Recognition and Artificial Intelligence: A Systematic Review (2014-2023) and Research Recommendations' (2024) 102 Information Fusion 102019. doi:10.1016/j.inffus.2023.102019

19. Kumar V and others, 'Understanding the Role of Artificial Intelligence in Personalized Engagement Marketing' (2019) 61(4) California Management Review 135. doi:10.1177/0008125619859317

20. Kurian N, 'AI's Empathy Gap: The Risks of Conversational Artificial Intelligence for Young Children's Well-Being and Key Ethical Considerations for Early Childhood Education and Care' (2023) 26(1) Contemporary Issues in Early Childhood 132. doi:10.1177/14639491231206004

21. Li Y and others, 'Developing Trustworthy Artificial Intelligence: Insights from Research on Interpersonal, Human-Automation, and Human-AI Trust' (2024) 15 Frontiers in Psychology 1382693. doi:10.3389/fpsyg.2024.1382693

22. Mantello P and others, 'Machines that Feel: Behavioral Determinants of Attitude Towards Affect Recognition Technology—Upgrading Technology Acceptance Theory with the Mind sponge Model' (2023) 10(1) Humanities and Social Sciences Communications 430. doi:10.1057/s41599-023-01837-1

23. McStay A, 'Emotional AI, Soft Biometrics and the Surveillance of Emotional Life: An Unusual Consensus on Privacy' (2020) 7(1) Big Data & Society. doi:10.1177/2053951720904386

24. McStay A, *Emotional AI: The Rise of Empathic Media* (SAGE Publications 2018) doi:10.4135/9781526451293

25. McStay A, 'Empathic Media and Advertising: Industry, Policy, Legal and Citizen Perspectives (The Case for Intimacy)' (2016) 3(2) Big Data & Society. doi:10.1177/2053951716666868

26. Menges L and Weber-Guskar E, 'Digital Emotion Detection, Privacy, and the Law' (2025) 38(2) Philosophy & Technology 1. doi:10.1007/s13347-025-00895-4

27. Nag PK, Bhagat A and Priya RV, 'Expanding AI's Role in Healthcare Applications: A Systematic Review of Emotional and Cognitive Analysis Techniques' [2025] IEEE Access. doi:10.1109/ACCESS.2025.3562131

28. Palmiotto F and González NM, 'Facial Recognition Technology, Democracy and Human Rights' (2023) 50 *Computer Law & Security Review* 105857. doi:10.1016/j.clsr.2023.105857

29. Pan B and others, 'A Review of Multimodal Emotion Recognition from Datasets, Preprocessing, Features, and Fusion Methods' (2023) 561 Neurocomputing 126866. doi:10.1016/j.neucom.2023.126866

30. Picard RW, *Affective Computing* (Technical Report no 321, MIT Media Laboratory Perceptual Computing Section 1995)

31. Podoletz L, 'We Have to Talk A bout Emotional AI and Crime' (2023) 38 AI & Society 1067. doi:10.1007/s00146-022-01435-w

32. Salvi del Pero A, Wyckoff P and Vourc'h A, *Using Artificial Intelligence in the Workplace: What are the Main Ethical Risks?* (Social, Employment and Migration Working Papers no 273, OECD 2022). doi:10.1787/840a2d9f-en

33. Siau KL and Wang W, 'Building Trust in Artificial Intelligence, Machine Learning, and Robotics' (2018) 31(2) Cutter Business Technology Journal 47

34. Thakkar A, Gupta A and De Sousa A, 'Artificial Intelligence in Positive Mental Health: A Narrative Review' (2024) 6 Frontiers in Digital Health 1280235. doi:10.3389/fdgth.2024.1280235

35. Varsha PS, 'How Can We Manage Biases in Artificial Intelligence Systems: A Systematic Literature Review' (2023) 3(1) International Journal of Information Management Data Insights 100165. doi:10.1016/j.jjimei.2023.100165

36. Velagaleti SB and others, 'Empathetic Algorithms: The Role of AI in Understanding and Enhancing Human Emotional Intelligence' (2024) 20(3) Journal of Electrical Systems 2051. doi:10.52783/jes.1806

37. Vistorte AOR and others, 'Integrating Artificial Intelligence to Assess Emotions in Learning Environments: A Systematic Literature Review' (2024) 15 Frontiers in Psychology 1387089. doi:10.3389/fpsyg.2024.1387089

38. Zepf S and others, 'Driver Emotion Recognition for Intelligent Vehicles: A Survey' (2020) 53(3) ACM Computing Surveys (CSUR) 64. doi:10.1145/338879

39. Zualkernan IA and others, 'Emotion Recognition Using Mobile Phones' (2017) 60 Computers & Electrical Engineering 1. doi:10.1016/j.compeleceng.2017.05.004

AUTHORS INFORMATION

Laroussi Chemlali*
PhD in Law, Associate Professor, Ajman University, College of Law, Ajman, United Arab Emirates
l.chemlali@ajman.ac.ae
https://orcid.org/0000-0002-4770-7121
Corresponding author, responsible for conceptualisation, methodology, writing – original draft, writing –review & editing, supervision, validation.

Leila Benseddik
PhD in Applied Linguistics, Assistant Professor, Canadian University of Dubai, Faculty of First Year, Dubai, United Arab Emirates
leila.benseddik@cud.ac.ae
https://orcid.org/0009-0000-4556-2961
Co-author, responsible for conceptualisation, methodology, writing – original draft, writing – review & editing, supervision.

Competing interests: No competing interests were disclosed. Any potential conflict of interest must be disclosed by authors.

Disclaimer: The authors declare that their opinion and views expressed in this manuscript are free of any impact of any organizations.

ACKNOWLEDGEMENTS

The authors wish to express their sincere appreciation to Ajman University for its financial support in covering the Article Processing Charges (APC) for this publication.

RIGHTS AND PERMISSIONS

EDITORS

ABOUT THIS ARTICLE

Cite this article
Chemlali L and Benseddik L, 'Privacy-By-Design in Emotion AI: Data Protection Frameworks and Compliance Strategies' (2025) 8(Spec) Access to Justice in Eastern Europe 288-311 <https://doi.org/10.33327/AJEE-18-8.S-r000153>

DOI: https://doi.org/10.33327/AJEE-18-8.S-r000153

Keywords: *privacy-by-design, emotion AI, data protection, ethics in AI, emotion recognition, user privacy.*

DETAILS FOR PUBLICATION

Date of submission: 21 Aug 2025
Date of acceptance: 02 Oct 2025
Online First publication: 04 Dec 2025
Last Published: 30 Dec 2025
Whether the manuscript was fast tracked? - No
Number of reviewer report submitted in first round: 2 reports
Number of revision rounds: 1 round with major revisions

Technical tools were used in the editorial process:
Plagiarism checks - Turnitin from iThenticate https://www.turnitin.com/products/ithenticate/
Scholastica for Peer Review https://scholasticahq.com/law-reviews

AI DISCLOSURE STATEMENT

The corresponding author confirms that this article was prepared with the assistance of AI tools. Specifically, ChatGPT (OpenAI) and Claude were used for language refinement during the drafting process. All content, arguments, and conclusions were generated independently and remain their sole responsibility. No AI tool was used for generating original research findings or analysis.

АНОТАЦІЯ УКРАЇНСЬКОЮ МОВОЮ

Оглядова стаття

КОНФІДЕНЦІЙНІСТЬ ЗА ЗАМОВЧУВАННЯМ В ЕМОЦІЙНОМУ ШІ: ЗАХИСТ ДАНИХ ТА СТРАТЕГІЇ ДОТРИМАННЯ ВИМОГ

Ларуссі Чемлалі* та Лейла Бенседдік

АНОТАЦІЯ

Вступ. *Швидкий розвиток емоційного штучного інтелекту (емоційний ШІ) створив значні можливості для інновацій у широкому спектрі галузей, зокрема в сфері охорони здоров'я, маркетингу та взаємодії людини з комп'ютером. Застосунки з емоційним ШІ, які обробляють, аналізують та реагують на людські емоції, здебільшого залежать від приватних персональних даних, що призводить до проблем із конфіденційністю та етичними нормами. Впровадження принципів конфіденційності за замовчуванням (PbD) у таких системах є важливим для протидії цим викликам та забезпечення відповідності змінам у правових системах. У цій статті розглядається взаємодія між PbD та емоційним ШІ, з наголосом на ризики для конфіденційності, що пов'язані зі збором та обробкою емоційних даних. Дослідження проводиться на тлі ширшого контексту розробки етичного ШІ. У роботі було підкреслено нагальну необхідність збалансувати технологічні інновації з надійним захистом конфіденційності.*

Методи. *У статті пропонується концептуальний юридичний аналіз взаємодії між концепцією конфіденційності за замовчуванням (PbD) та емоційним ШІ у сучасних системах захисту даних. У ній використовується комплексний огляд первинних джерел, зокрема GDPR ЄС, Акт ЄС про ШІ, судову практику Суду ЄС та ЄСПЛ, Рекомендації щодо захисту персональних даних, а також вторинні джерела, такі як наукові праці та книги. Дослідження структуровано таким чином: спочатку подано огляд емоційного ШІ, його застосування, а також проблеми конфіденційності, які він спричиняє. Далі йде розгляд наявних режимів захисту даних та того, як їх можна перенести на системи емоційного ШІ. Після чого увага зосереджується на*

фундаментальних принципах PbD, а також вивчається, як їх можна застосовувати під час розробки та впровадження технологій емоційного ШІ.

***Результати та висновки**. Аналіз демонструє, що впровадження принципів PbD у системи емоційного ШІ є важливим, а не просто корисним для того, щоб захистити конфіденційність користувачів та забезпечити дотримання законодавства. Правильно впроваджені системи PbD мають три важливі переваги: підвищену прозорість системи, кращі механізми підзвітності та більший контроль користувачів над власними даними. Ці висновки роблять значний внесок у теоретичні основи відповідального проєктування штучного інтелекту, пропонуючи практичні рекомендації для організацій, що впроваджують системи емоційного ШІ. Зрештою, дослідження представляє чітку модель для розробників та організацій, яка допоможе їм успішно скористатися конвергенцією технологій емоційного інтелекту та нормативних вимог щодо конфіденційності.*

***Ключові слова.** конфіденційність за замовчуванням, емоційний ШІ, захист даних, етика у ШІ, розпізнавання емоцій, конфіденційність користувачів.*

ABSTRACT IN ARABIC

مقالة مراجعة

الخصوصية المدمجة في التصميم ضمن تقنيات الذكاء الاصطناعي العاطفي: أطر حماية البيانات واستراتيجيات الامتثال

العروسي الشملالي* و ليلى بن صديق

الملخص

***خلفية الدراسة:** أسهم التطور السريع في مجال الذكاء الاصطناعي العاطفي في فتح مجالات واسعة للابتكار في قطاعات متعددة تشمل الرعاية الصحية والتسويق والتفاعل بين الإنسان والآلة. وتعتمد تطبيقات هذا النوع من الذكاء الاصطناعي، القائم على تحليل الانفعالات البشرية ومعالجتها والتفاعل معها، على بيانات شخصية شديدة الحساسية، الأمر الذي يثير تحديات جوهرية تتعلق بالخصوصية وقضايا أخلاقية متنامية. ويعد تبني مبادئ الخصوصية المدمجة في التصميم داخل هذه الأنظمة خطوة أساسية لمواجهة هذه التحديات وضمان الامتثال للأطر القانونية المتغيرة. وتتناول هذه الدراسة العلاقة بين الخصوصية المدمجة في التصميم والذكاء الاصطناعي العاطفي، مع التركيز على المخاطر التي*

تهدد خصوصية الأفراد نتيجة جمع البيانات الانفعالية ومعالجتها. وتأتي الدراسة ضمن سياق أشمل يركز على تطوير ذكاء اصطناعي أخلاقي، مؤكدة الحاجة الملحة إلى المواءمة بين التقدم التقني وتوفير حماية قوية وفعّالة لخصوصية المستخدمين.

المنهجية: *يقدم هذا البحث تحليلًا قانونيًا مفاهيميًا للتقاطع بين الخصوصية المدمجة في التصميم والذكاء الاصطناعي العاطفي ضمن أطر حماية البيانات الحديثة. ويستند إلى مراجعة شاملة للمصادر الأولية، بما في ذلك اللائحة العامة لحماية البيانات في الاتحاد الأوروبي، وقانون الذكاء الاصطناعي الأوروبي، وأحكام محكمتي العدل الأوروبية وحقوق الإنسان الأوروبية، إلى جانب الإرشادات الصادرة عن هيئات حماية البيانات، فضلًا عن المصادر الثانوية مثل الدراسات الأكاديمية والكتب المتخصصة. وقد نُظمت المناقشة بحيث تبدأ بعرض شامل لمفهوم الذكاء الاصطناعي العاطفي وتطبيقاته والمخاوف الفردية المتعلقة بالخصوصية التي يثيرها. ويتبع ذلك تناولٌ للأطر الحالية لحماية البيانات وسبل تكييفها لتناسب أنظمة الذكاء الاصطناعي العاطفي. ثم تنتقل الدراسة إلى مناقشة المبادئ الأساسية للخصوصية المدمجة في التصميم، مع بحث كيفية تطبيق هذه المبادئ عند تطوير ونشر تقنيات الذكاء الاصطناعي العاطفي.*

النتائج والاستنتاجات: *يبيّن التحليل أن تطبيق مبادئ الخصوصية المدمجة في التصميم داخل أنظمة الذكاء الاصطناعي العاطفي ليس خيارًا إضافيًا، بل ضرورة أساسية لحماية خصوصية المستخدمين وضمان الامتثال القانوني. وعندما تُنفذ هذه المبادئ بصورة سليمة، فإنها توفر ثلاث فوائد محورية تتمثل في تعزيز شفافية النظام، وتقوية آليات المساءلة، وتوسيع قدرة الأفراد على التحكم في بياناتهم الشخصية. وتسهم هذه النتائج إسهامًا مهمًا في ترسيخ الأسس النظرية لتصميم ذكاء اصطناعي مسؤول، كما تقدم إرشادات عملية قابلة للتطبيق للجهات التي تعمل على تطوير ونشر تقنيات الذكاء الاصطناعي العاطفي. وتختتم الدراسة بطرح نموذج واضح يمكّن المطورين والمؤسسات من مواكبة التلاقي المتسارع بين تقنيات الذكاء العاطفي وتطورات التنظيمات المرتبطة بالخصوصية.*

Access to Justice in Eastern Europe
ISSN 2663-0575 (Print)
ISSN 2663-0583 (Online)
http://ajee-journal.com

OPINION ARTICLE

Opinion Article

CHATGPT IN THE DOCK: REFLECTIONS ON THE FUTURE OF CRIMINAL LIABILITY

Raed S A Faqir

DOI:
https://doi.org/10.33327/AJEE-18-8.S-o000156

Date of submission: 19 Aug 2025
Date of acceptance: 30 Oct 2025
Date of Publication: 30 Dec 2025

Disclaimer:

The author declares that his opinion and views expressed in this manuscript are free of any impact of any organizations.

ABSTRACT

Background: *This study examines the legal challenges posed by generative AI. It highlights the limitations of traditional criminal liability frameworks in addressing harm caused by AI outputs. The research explores new models of liability to ensure accountability while protecting individual rights in the age of intelligent machines.*

Generative AI, exemplified by ChatGPT, has evolved from a mere computational tool into a cognitive agent capable of content creation, problem-solving, and decision-making. This evolution challenges traditional criminal law frameworks, raising complex questions about the attribution of Liability when AI-generated outputs result in harm or criminal conduct. The study explores these dilemmas, focusing on the shortcomings of conventional concepts of criminal liability and exploring the need for new legal paradigms.

Methods: *The research employs a descriptive-analytical and comparative methodology. It analyses national and international legislation, legal principles, and contemporary jurisprudence, with a focus on the European Artificial Intelligence Act (2024) as a model. The study examines AI's*

autonomous capabilities, the opacity of algorithmic decision-making, and the challenges of establishing causal links between AI actions and resulting harms. Case studies are used to explore potential liability models, including preventive liability and the concept of an "artificial actor."

Results and Conclusions: *The study finds that traditional frameworks of criminal accountability are inadequate for AI systems like ChatGPT, given their partial autonomy and algorithmic complexity. It highlights the potential for expanding liability to developers, operators, and users, and the necessity of flexible legal models that combine preventive, administrative, and criminal measures. The research underscores the importance of integrating legal innovation with technological oversight to safeguard individual rights while maintaining the deterrent and protective functions of criminal law.*

1 INTRODUCTION

Generative Artificial Intelligence, led today by models such as ChatGPT, represents a pivotal milestone in the trajectory of global technological transformation. It is no longer merely an executive tool but has become a cognitive agent actively participating in content creation, problem-solving, and decision-making. This shift is clear in the growing reliance on AI systems across multiple fields—including education, healthcare and law—posing unprecedented challenges to legal systems, particularly the criminal justice system. The system now faces uncertainty in identifying the "actor" in cases involving harm or criminalised conduct resulting from AI intervention.

From this standpoint, the present study, titled *ChatGPT in the Dock: Reflections on the Future of Criminal Liability*, aims to analyse the legal dilemmas posed by this model, one of the most prominent and controversial manifestations of generative artificial intelligence. Its primary objectives are: (1) to expose the shortcomings of traditional concepts of criminal liability in light of AI intervention; (2) to analyse the complex structure of liability resulting from the actions and outputs of systems like ChatGPT; and (3) to propose restructuring legal attribution rules in a manner that ensures the protection of individual rights and the effectiveness of criminal deterrence in the age of intelligent machines.

Upon AI autonomy, the study takes a two-pronged approach: theoretically, it examines criminal concepts such as intent, causation, and actor liability; practically, it examines comparative legal models, particularly the 2024 European AI Act, and their applicability to cases involving harmful or deceptive AI use, like ChatGPT. The study is grounded in a central hypothesis: current frameworks of criminal accountability are incapable of comprehending the outputs of partially autonomous non-human entities. This necessitates the development of more flexible legal models, such as collective or virtual liability, or even the establishment of a new concept: the "artificial actor."

The core legal questions this study addresses include: Who bears criminal liability when ChatGPT use results in harmful or criminal outputs? Should liability be limited to the end user, or should it extend to developers, operators, and designers? To what extent can ChatGPT be considered an independent actor contributing to the criminal outcome? What is the optimal legislative pathway to bridge the legal gap caused by the ambiguity of intent and discernment in AI-generated actions? These questions are not merely theoretical; they lie at the heart of the challenges facing criminal justice in the coming decades. Exploring them constitutes a legal and strategic necessity for ensuring the sustainability of the judicial system in a rapidly evolving digital environment.

This study on criminal liability for AI systems like ChatGPT addresses the shortcomings of existing legal frameworks, identifies the parties liable, and the challenges in assessing intent. It examines legitimate risks, regulatory gaps, and the difficulties of applying traditional criminal law to autonomous AI behaviour. Finally, it proposes future-oriented solutions to change liability frameworks, such as corporate responsibility and systemic accountability.

2 LITERATURE REVIEW

The scientific literature indicates that criminal accountability for artificial intelligence systems, particularly complex software such as ChatGPT, faces fundamental challenges related to the nature of algorithmic actions and the opacity of decision-making processes.[1] Many researchers point out that AI operates through intricate algorithmic networks beyond human control, making it more difficult to attribute any resulting harm to conventional ideas of fault or intent.[2] Traditional criminal liability, which is based on direct human action, is being reexamined amid the crisis of causal attribution.[3]

In the same context, recent legal studies have addressed the crisis of lack of control over the technical risks posed by AI, illustrating that traditional laws are incapable of regulating these new technological risks within conventional liability frameworks.[4] To

1 Alejo José G Sison and others, 'ChatGPT: More than a "Weapon of Mass Deception" Ethical Challenges and Responses from the Human-Centered Artificial Intelligence (HCAI) Perspective' (2024) 40(17) International Journal of Human-Computer Interaction 4853. doi:10.1080/10447318.2023.2225931.

2 PR Biju and O Gayathri, 'Algorithmic Solutions, Subjectivity and Decision Errors: A Study of AI Accountability' (2025) 27(5) Digital Policy, Regulation and Governance 523. doi:10.1108/DPRG-05-2024-0090.

3 Marcelo Ferrante, 'Causation in Criminal Responsibility' (2008) 11(3) New Criminal Law Review 470. doi:10.1525/nclr.2008.11.3.470.

4 Benjamin Cheatham, Kia Javanmardian and Hamid Samandari, 'Confronting the Risks of Artificial Intelligence' (2019) 2 McKinsey Quarterly 8 <https://www.mckinsey.com/capabilities/quantumblack/our-insights/confronting-the-risks-of-artificial-intelligence> accessed 10 August 2025.

address risks posed by software like ChatGPT, the legal landscape is moving toward preventive and abstract liability models, emphasising producers' accountability for technical safety measures rather than for direct harm.[5]

Moreover, the literature highlights legislative and practical challenges stemming from the exclusion of intelligent software from traditional product laws, particularly those that define a product solely as a physical entity, thereby obstructing effective accountability for such systems.[6] In this regard, the European Artificial Intelligence Act emerges as an advanced model imposing deterrent sanctions on violations and promoting enhanced transparency and disclosure of AI-related risks.[7] Recent studies also recommend integrating administrative oversight with criminal liability, thereby opening new horizons for managing technological risks through an integrated and evolving legal framework that safeguards fundamental rights without impeding technological innovation.[8]

3 METHODOLOGY

This study aims to analyse the legal framework governing criminal liability for generative artificial intelligence systems, focusing on the ChatGPT model as a contemporary practical example that reflects the legal and technical challenges in this field. To achieve this, the study employs a descriptive-analytical approach that examines national and international legislative texts and relevant legal principles, and reviews contemporary jurisprudential and legal literature on the nature of criminal liability amid the rapid development of AI technologies. Additionally, the study employs a comparative legal method, which systematically identifies and investigates specific areas where legal systems diverge and converge. The comparative criteria include factors such as the degree of criminal liability, regulatory protection, enforcement tactics, and ethical accountability standards. Using this framework, the study contrasts advanced European legislation, particularly the Artificial

5 ibid 9.

6 Omena Akpobome, 'The Impact of Emerging Technologies on Legal Frameworks: A Model for Adaptive Regulation' (2024) 5(7) International Journal of Research Publication and Reviews 5049. doi:10.55248/gengpi.5.1024.3012.

7 M Navaneeth, 'The Need for A Global Regulatory Framework for Artificial Intelligence: Implications of the European Union European Union Artificial Intelligence Act 2024' (Master's thesis, National University of Advanced Legal Studies 2024) 62-77; Mohammed Salem Alneyadi and others, 'The Crime of Electronic Blackmail in the Emirati Law' (2022 International Arab Conference on Information Technology (ACIT), Abu Dhabi, UAE, 22-24 November 2022). doi:10.1109/ACIT57182.2022.9994165.

8 Jennifer Kuzma and others, 'An Integrated Approach to Oversight Assessment for Emerging Technologies' in Gary E Marchant and Wendell Wallach (eds), *Emerging Technologies: Ethics, Law and Governance* (Routledge 2020) 1199. doi:10.1111/j.1539-6924.2008.01086.x.

Intelligence Act (AI Act),[9] with local legislative systems that are still in their infancy in addressing emerging technological challenges.[10]

The study, employing a critical approach, assesses the effectiveness of current legal frameworks in addressing AI-related harms, emphasising the challenges of establishing traditional causal links between actions and outcomes in algorithmic contexts. It explores the potential adoption of new liability models centred on preventive liability and criminal negligence, using case studies to illustrate how laws can evolve to balance societal protection with the promotion of technological innovation.

4 CRIMINAL LIABILITY DEFINITION FOR ARTIFICIAL MINDS AND CHATGPT

4.1. A Mind Without a Body: Who Prosecutes ChatGPT?

The phenomenon of ChatGPT vividly exemplifies the profound complexities artificial intelligence introduces into the criminal legal system.[11] This advanced linguistic system does not merely process data, but generates textual decisions that interact with humans and influence their cognitive, social, and legal realities.[12] When the generated text becomes capable of shaping convictions or guiding decisions, we are no longer dealing with a mere silent technical tool but a virtual mind without a body—one that redefines legal agency.[13] This raises fundamental questions about the nature and legal classification of artificial intelligence, especially in the absence of a unified definition within legal systems.[14] Traditional criminal models of intent, perpetrator identification, and liability must be reevaluated as AI's legal identity straddles the line between a human-controlled tool and an autonomous decision-maker.[15]

9 Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 'Laying Down Harmonised Rules on Artificial Intelligence and Amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828' (Artificial Intelligence Act - IA Act) [2024] OJ L 1689 <https://eur-lex.europa.eu/eli/reg/2024/1689/oj> accessed 10 August 2025.

10 Navaneeth (n 7) 62-77.

11 Christiaan Mineur, 'Autonomous AI Technology and the Evolution of Legal Personhood in Criminal Law' (Master's thesis, University College Tilburg 2024) 4.

12 Iman M Al-Uqdah, 'Criminal Liability for Artificial Intelligence Application Crimes' (2025) 153 Journal of Law and Jurisprudence 39.

13 Mineur (n 11) 8.

14 Maxi Scherer, 'Artificial Intelligence and Legal Decision-Making: The Wide Open?' (2019) 36(5) Journal of international arbitration 541. doi:10.54648/joia2019028.

15 Jacob Turner, 'Legal Personality for AI' in Jacob Turner, *Robot Rules: Regulating Artificial Intelligence* (Springer 2018) 175. doi:10.1007/978-3-319-96235-1_5.

When we ask, *"Who prosecutes ChatGPT?"*, we pierce through the veil of classical law and enter an unprecedented legal space.[16] The European Commission's 2021 proposal,[17] despite its effort to define AI broadly as systems capable of "generating outputs that affect the environment," fails to address the core dilemma: how to distinguish between basic AI and those complex generative systems that produce socially and legally impactful texts—like ChatGPT.[18] The challenge lies in three central characteristics: autonomy, interactivity, and opacity.[19] The term "autonomy" describes AI's capacity to act without direct oversight, not its intention. While opacity reflects the "black box" nature of its unpredictable algorithms, interactivity arises from daily user engagement.[20] These traits complicate the attribution of criminal liability, as the lines blur between programming error and human intention, between spontaneous output and directed decision.[21]

The dilemma posed by ChatGPT goes beyond legal debate into deep philosophical and ethical territory, shaking the foundations of traditional criminal concepts.[22] The absence of premeditation, the unpredictability of outputs, and the difficulty in identifying a clear actor, be it the developer, the user, or the owning company, reshapes the question of criminal liability.[23] It transforms it from a simple binary framework into a multi-layer network.[24] The traditional legal system, built on the formula "actor–victim–harm," is no longer adequate to encompass intelligent entities that commit crimes not in conventional ways but through knowledge flows open to interpretation.[25]

16 Amirreza Ahkami, 'AI and The European Union's Approach to Data Protection: The Case of Chat GPT' (Master's thesis, University of Padova 2024) 33-4.

17 European Commission, 'Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts' (COM/2021/206 final, 21 April 2021) <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206> accessed 10 August 2025.

18 Navaneeth (n 7) 62-77.

19 Bram Vaassen, 'AI, Opacity, and Personal Autonomy' (2022) 35(4) Philosophy & Technology 89. doi:10.1007/s13347-022-00577-5.

20 Youliang Yuan and others, 'Does ChatGPT Know that it Does Not Know? Evaluating The Black-Box Calibration of Chatgpt' (2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), Torino, Italia, 20-25 May 2024) 5191. See also, Jingyu Wang and others, 'Network Meets ChatGPT: Intent Autonomous Management, Control and Operation' (2023) 8(3) Journal of Communications and Information Networks 340. doi:10.23919/JCIN.2023.10272352.

21 Briony Blackmore, 'Looking Beyond Blame and Praise: Analyzing Moral Responsibility in the Development and Deployment of AI Systems' (PhD thesis, University of Otago 2023) 12.

22 Al-Uqdah (n 12) 56.

23 Scherer (n 14) 542.

24 Xin Chen, 'Research on the Application of Intelligent ChatGPT in Computer Intelligent Computing System' (2023 IEEE 3rd International Conference on Data Science and Computer Application (ICDSCA), Dalian, China, 27-29 October 2023) 985. doi:10.1109/icdsca59871.2023.10392475.

25 Turner (n 15) 174.

Therefore, there is a pressing need for a new legal model based on "distributed shared accountability" and transparent oversight mechanisms that allow for tracing algorithmic decisions and assessing their legality and impact.[26] Ultimately, prosecuting a "mind without a body" is not a metaphor—it is a tangible necessity that demands an epistemological revolution in our understanding of law and a careful balance between technological innovation and legal protection of rights and freedoms in the digital age.[27]

4.2. When ChatGPT Speaks Without Criminal Intent: Can It Be Prosecuted?

We are not here to explore the intricate technical workings that give ChatGPT its structure, as these have been thoroughly explained by specialists and go beyond the concerns of criminal jurisprudence.[28] Even in cases where there is no criminal intent or will, what matters is how AI-generated linguistic outputs affect the legal system.[29] The machine learning model does not operate on explicit logical rules; rather, it follows a probabilistic inductive approach, extracting linguistic patterns from billions of examples without "understanding" them. ChatGPT neither knows the truth nor intends to lie, yet it can generate harmful, misleading, or inflammatory content.[30] AI thus generates "speech without reason" and "action without intent," undermining traditional legal theories that attribute criminal liability exclusively to human consciousness and rational awareness.[31]

When ChatGPT generates illegal or criminally consequential content, the challenge of identifying the responsible actor emerges. Is it the model itself? The developers? The owning company? Or the users? Like a "blind painter," the model uses probabilistic estimates that change as its inputs change to create linguistic portraits without knowing why or for what purpose.[32] To maximise flexibility and generative capacity rather than the logical consistency required for legal liability, its architecture is deliberately opaque.[33] Under the new legal concept of distributed liability, traditional actors cannot be held

26 Kuzma and others (n 8) 1198.

27 Akpobome (n 6) 5050.

28 Kalliopi Terzidou, 'Generative AI for the Legal Profession: Facing the Implications of the Use of ChatGPT Through an Intradisciplinary Approach' (*Media Laws*, 8 September 2023) 4 <https://www.medialaws.eu/generative-ai-for-the-legal-profession-facing-the-implications-of-the-use-of-chatgpt-through-an-intradisciplinary-approach/> accessed 10 August 2025.

29 Aslihan Asil and Thomas G Wollmann, 'Can Machines Commit Crimes Under US Antitrust Laws?' (2024) 3(1) The University of Chicago Business Law Review 6 <https://businesslawreview.uchicago.edu/print-archive/can-machines-commit-crimes-under-us-antitrust-laws > accessed 10 August 2025.

30 Terzidou (n 28) 2.

31 Lawrence B Solum, 'Legal Personhood for Artificial Intelligences' (1992) 70 North Carolina Law Review 1272-3.

32 Sarah Muller, 'Visual Silence in the Language Portrait: Analyzing young People's Representations of their Linguistic Repertoires' (2022) 25(10) International Journal of Bilingual Education and Bilingualism 3646. doi:10.1080/13670050.2022.2072170.

33 Blackmore (n 21) 44.

accountable; instead, accountability must be traced across multiple stakeholders, from user interfaces to training environments.[34]

Even though ChatGPT does not aim to break the law or incite, its reliance on linguistic patterns can yield results with significant ethical or legal implications. The legal challenge lies here: how do we assign accountability to a system that lacks will, yet produces effects? The answer requires moving beyond traditional liability frameworks by developing new rules that hold developers and operators accountable and enforce proactive oversight of inputs and algorithms. The legal focus must shift from "the actor's intent" to "design and operational responsibility," and from the "criminal mind" to a system of "digital governance." This model does not err because it chooses to, but because it lacks the capacity to distinguish right from wrong—necessitating legislation that redefines the relationship between technology and accountability under a new logic.[35]

ChatGPT stands at the threshold of artificial consciousness, in a legal grey zone that criminal justice systems are not yet prepared to handle. It does not think or comprehend, but it generates discourse that simulates thought. It is neither a traditional actor nor a mere tool—it is a linguistic entity that challenges settled legal classifications. In this sense, artificial intelligence functions more as a mirror exposing the inadequacies of our laws than as a standalone problem. The challenge lies not in how "intelligent" it is, but in how legally "prepared" we are to incorporate it into our network of criminal concepts. A reevaluation of crime and punishment in which actors may be nonhuman, and liability arises from error, probability, or unanticipated consequences rather than conscious intent, would result from failing to act, risking the prosecution of algorithms for unintended outputs.

4.3. Attributing Criminal Liability in the Age of Intelligent Machines

Artificial intelligence systems pose a genuine challenge to the traditional criminal liability framework, which is built upon the pillars of actus reus (the act), mens rea (intent), and will.[36] These intelligent entities are neither human nor self-aware nor criminally intent; rather, they are digital tools that generate unpredictable behaviours that are difficult to foresee accurately.[37] As reliance on these complex systems—operating on probabilistic rather than explicit logical bases—increases, a central legal question arises: How can liability for harm caused by these systems be assigned when humans lack full control over their behaviour? ChatGPT and similar language models do not rely on true understanding. Still, on intricate statistical patterns they neither comprehend nor can

34 Dirk A Zetzsche, Ross P Buckley and Douglas W Arner, 'The Distributed Liability of Distributed Ledgers: Legal Risks of Blockchain' (2018) 4 University of Illinois Law Review 1386. doi:10.2139/ssrn.3018214.

35 Akpobome (n 6) 5050.

36 Mineur (n 11) 22.

37 Asil and Wollmann (n 29) 20.

explain, making prosecution for harmful actions impossible under traditional criminal law concepts.[38] Hence, our legal system faces a profound crisis in how to classify these non-living entities that produce real-world effects and in defining a legal framework balancing technological progress with justice.[39]

Confronting this dilemma, international approaches, especially within the European Union, have emphasised the importance of strict civil liability as a practical mechanism to protect victims.[40] Directive 85/374/EEC on defective products is a law that holds manufacturers liable for damages without needing proof of fault or intent.[41] Traditional concepts of "defect" and "causality" are blurred by the complexity of AI since intelligent systems are dynamic entities with probabilistic behaviours rather than traditional products.[42] This prompted the European Commission in 2022 to propose comprehensive legal updates addressing "smart products".[43] These updates aim to broaden legal protection and shift the burden of proof onto producers by presuming a link between defect and damage automatically, and imposing economic responsibility on producers for risks associated with these systems—even when their behaviour is unpredictable or unforeseeable.[44] This reflects a fundamental shift in liability philosophy—from focusing on the actor's intent to ensuring effective compensation for victims regardless of the actor's awareness.[45]

Since AI is an unconscious entity lacking intent or will, the law must transcend traditional concepts grounded in these elements and develop a hybrid legal framework combining strict civil responsibility, regulatory liability, and supervisory oversight.[46] Producers or developers oversee the implementation of safety precautions and ensure clear standards, aiming to prevent harm and provide victims with straightforward compensation.[47] Instead of prosecuting algorithms without understanding how they work, the AI approach

38 Mineur (n 11) 19.
39 Scherer (n 14) 542.
40 Ahkami (n 16) 39.
41 Council Directive 85/374/EEC of 25 July 1985 'On the Approximation of the Laws, Regulations and Administrative Provisions of the Member States Concerning Liability for Defective Products' [1985] OJ L 210/29; Fidelma White, 'Directive 85/374/EEC Concerning Liability for Defective Products: In the Name of Harmonisation, the Internal Market and Consumer Protection' in Paula Giliker (ed), *Research Handbook on EU Tort Law* (Edward Elgar 2017) 128. doi:10.4337/9781785365720.00013.
42 Keith Darlington, 'Aspects of Intelligent Systems Explanation' (2013) 1(2) Universal Journal of Control and Automation 47. doi:10.13189/ujca.2013.010204.
43 European Commission, 'Proposal for a Regulation of the European Parliament and of the Council on Horizontal Cybersecurity Requirements for Products with Digital Elements and Amending Regulation (EU) 2019/1020' (COM/2022/454 final, 15 September 2022) <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52022PC0454> accessed 10 August 2025; Ahkami (n 16) 39.
44 Navaneeth (7) 62-77.
45 Cheatham, Javanmardian and Samandari (n 4) 9.
46 Kuzma and others (n 8) 1197.
47 Oscar Oviedo-Trespalacios and others, 'The Risks of Using ChatGPT to Obtain Common Safety-Related Information and Advice' (2023) 167 Safety Science 106244. doi:10.1016/j.ssci.2023.106244.

emphasises accountability for system design and operation, prevention, and compensation.[48] Without renewing our legal frameworks, we risk a future in which algorithms are prosecuted for unintended or unforeseen actions, opening a dangerous legal vacuum that threatens the application of justice.[49]

5 CHALLENGES IN ASSESSING THE CRIMINAL LIABILITY OF ARTIFICIAL INTELLIGENCE

Conventional criminal law faces unprecedented challenges because of the rise of ChatGPT and other generative AI systems. At the core of this disruption is the causality dilemma: it is impossible to use conventional legal methods to link AI actions to outcomes. The fact that AI operates without conscious intent, which contradicts conventional notions of guilt, exacerbates the *mens rea* dilemma. When taken as a whole, these crises highlight the pressing need to reconsider liability assessment in the age of intelligent machines.

5.1. The Crisis of Criminal Law in the Era of ChatGPT

The digital revolution is causing a major crisis for criminal law, which is still confined to traditional frameworks based on human consciousness, intent, and choice.[50] The criminal system is governed by well-established principles such as legality, personal blame, and the presumption of innocence, designed to address human actions with clear motives. Still, it is ill-equipped to accommodate acts generated by generative AI systems like ChatGPT.[51] Determining criminal liability in systems with self-learning capabilities and programmers' autonomy is challenging because of unpredictable behaviours that cannot be traced to specific human actors.[52] This clash represents the "shock of modernity" described by the Italian thinker Federico Stella, in which traditional criminal concepts such as intent and awareness lose their relevance when confronted with technology characterised by behavioural ambiguity and unpredictability.[53]

AI-induced harm challenges the evidentiary and proof mechanisms of criminal law because algorithmic biases or training errors make it more difficult to assign blame and to establish a

48 Sonia K Katyal, 'Private Accountability in the Age of Artificial Intelligence' (2019) 66(1) UCLA Law Review 90.

49 Rebecca Crootof, '"Cyborg Justice" and the Risk of Technological-Legal Lock-In' (2019) 119(7) Columbia Law Review 249.

50 Sergio Carrera, Valsamis Mitsilegas and Marco Stefan, *Criminal Justice, Fundamental Rights and the Rule of Law in the Digital Age: Report of a CEPS and QMUL Task Force* (CEPS 2021). 62.

51 Blackmore (n 21) 44.

52 Monika Simmler, 'Responsibility Gap or Responsibility Shift? The Attribution of Criminal Responsibility in Human-Machine Interaction' (2024) 27(6) Information, Communication & Society 1145. doi:10.1080/1369118X.2023.2239895.

53 Solum (n 31) 1274.

direct causal relationship between the action and the harm.[54] In a participatory, multi-agent technological environment, personal liability becomes unclear as multiple parties' acts and omissions intertwine, weakening courts' certainty and threatening the principle of predictability of outcomes.[55] For instance, output generated by ChatGPT may be used to incite, defame, or commit fraud, but the absence of direct human intent complicates criminal accountability and exposes the limitations of current legal tools to address these complexities.[56]

The natural outcome of this crisis is a clear and troubling legal vacuum, where current criminal law lacks explicit and comprehensive rules addressing acts resulting from AI that lack awareness or will.[57] These vacuum places justice in a predicament: AI itself cannot bear criminal liability, nor can the traditional responsible human—whether developer or user—be easily held liable due to algorithmic complexity and the difficulty of proving fault and intent.[58] Consequently, today's digital reality demands a rethinking of the very definitions of crime and the principles of liability, opening the door to profound legal and philosophical debates about the limits of criminal law and how to develop a regulatory framework that balances societal protection from AI risks with encouraging innovation without threatening legal security and justice.[59]

5.2. The Causality Dilemma in the Era of Generative Artificial Intelligence

Generative AI systems like ChatGPT challenge traditional legal notions of causality.[60] Criminal law's deterministic model linking human acts to outcomes struggles against AI's probabilistic algorithms,[61] which produce unexpected outcomes without a human actor,[62] undermining classical proof and responsibility frameworks.[63]

54 Henrique Manuel Gil Martins, 'Liability Implications of Artificial Intelligence use in Health: Fault and Risk in Public Sector Healthcare' (Master's thesis, Universidade Catolica Portuguesa 2020) 31-2; Sander Beckers, Hana Chockler and Joseph Halpern, 'A Causal Analysis of Harm' (Advances in Neural Information Processing Systems 35: 36th Conference on Neural Information Processing Systems (NeurIPS 2022), 28 November - 9 December 2022, New Orleans, Louisiana, USA) 2368.

55 Oren Griffiths and Anna Thorwart, 'Effects of Outcome Predictability on Human Learning' (2017) 8 Frontiers in Psychology 514. doi:10.3389/fpsyg.2017.00511.

56 Akpobome (n 6) 5051.

57 Solum (n 31) 1273-4; Zetzsche, Buckley and Arner (n 34) 1286.

58 Nora Osmani, 'The Complexity of Criminal Liability of AI Systems' (2020) 14(1) Masaryk University Journal of Law and Technology 59. doi:10.5817/mujlt2020-1-3.

59 Cheatham, Javanmardian and Samandari (n 4) 8.

60 Blackmore (n 21) 45.

61 Emad H Atiq, 'How Folk Beliefs about Free Will Influence Sentencing: A New Target for the Neuro-Determinist Critics of Criminal Law' (2013) 16(3) New Criminal Law Review 449. doi:10.1525/nclr.2013.16.3.449.

62 Douglas C Youvan, 'Reconciling the Probabilistic and Deterministic: Exploring Complexity, Emergence, and Uncertainty in Nature, AI, and Human Cognition' (Research Gate, October 2024) 6. doi:10.13140/RG.2.2.16020.10880.

63 ibid 7-8.

It is challenging to prove a causal link between the actions of generative AI systems and the outcomes they generate due to their technical complexity. This puts conventional legal reasoning to the test.[64] In the past, courts used instruments like digital "black boxes" that capture event data to piece together the sequence of an incident and identify its cause.[65] The algorithm's internal decision-making processes, however, are based on complex, nonlinear probabilistic models, and these tools only provide preliminary indications.[66] Experts and judges are unable to conclusively determine whether an error with ChatGPT results from a programming error, bias in the training data, or user behaviour.[67] Such ambiguities make it more difficult to establish clear causation, particularly in cases of harmful content generation, and complicate the application of conventional principles that rely on a clear connection between action and result.[68]

Criminal law faces an epistemic crisis due to this causality conundrum, compelling a reexamination of its central concepts of intent, causation, and liability.[69] The nomological-deductive model that judicial systems employ excludes other possible explanations and necessitates precise scientific law and a logical causal relationship between an action and its result.[70] Generative AI undermines this model, as even system developers cannot precisely identify the causes of the outcomes they produce, rendering actions into multidimensional probabilistic outcomes.[71] Legal systems must accept "flexible causality", place blame on programmers, users, and AI systems, and update evidentiary standards to account for social and technical context to ensure criminal justice.[72] Law must adapt to this new reality, avoiding confinement within rigid models that fail to accommodate the algorithmic revolution, thus preserving the essence of justice and individual rights in the age of artificial intelligence.[73]

64 Beckers, Chockler and Halpern (n 54) 2368.
65 Yuan and others (n 20) 5192.
66 Wang-Ji Yan and others, 'Navigating Uncertainties in Machine Learning for Structural Dynamics: A Comprehensive Review of Probabilistic and Non-Probabilistic Approaches in Forward and Inverse Problems' (arXiv:2408.08629, 16 August 2024) 16. doi:10.48550/arXiv.2408.08629.
67 Youvan (n 62) 6.
68 Osmani (n 58) 63.
69 Blackmore (n 21) 45.
70 Rocco Neri, 'Judging Beyond any Reasonable Doubt: A Logic and Epistemological Rule' (2024) 7 Quaestio Facti: Revista internacional sobre razonamiento probatorio 49-50. doi:10.33115/udg_bib/qf.i7.23028.
71 Atiq (n 61) 449.
72 Osmani (n 58) 63.
73 Beckers, Chockler and Halpern (n 54) 2371.

5.3. ChatGPT and the Mens Rea Dilemma

Amid the rapid development of generative artificial intelligence, the ChatGPT model has emerged as an icon of digital transformation in natural language processing.[74] However, this model embodies a fundamental legal crisis that touches the core of causal proof—a foundational element in determining criminal liability. Traditional legal concepts rely on a clear and direct causal relationship between a human act and a harmful result, based on a nomological model that infers harm from a specific act carried out with awareness and intent.[75] By contrast, ChatGPT, as a complex algorithmic system operating through millions of probabilities, does not produce actions with intent or purpose. Its outputs are based on data and deep learning, not on conscious decisions.[76]

For example, if ChatGPT generates false or inciting content that causes moral or material damage, a complex question arises: can it be proven that the model's output directly caused these outcomes? Or is the human user who deployed the content responsible? In such cases, traditional legal rules requiring the attribution of harm to a conscious being become difficult to apply when faced with the opacity of an "artificial mind" that possesses neither will nor intention.[77]

The mental element (*mens rea*) is a cornerstone of criminal liability, typically manifested in intent or negligence.[78] However, when discussing ChatGPT, these concepts begin to dissolve, as the model has no will or consciousness and thus cannot possess intent like a human actor.[79] This raises the question of whether producers or developers are exonerated. Not necessarily, as existing doctrines provide alternative bases for attribution, the law might invoke the notion of *probable intent*, holding developers or users accountable if they could reasonably foresee the harm as a potential consequence of using the model—even in the absence of malicious intent.[80]

Suppose a developer or company releases a version of ChatGPT capable of generating inciting or offensive content, knowing that it could be misused, yet fails to implement sufficient preventive measures. Here, *probable intent* is realised, as they are expected to foresee the risk and prevent it.[81] Conversely, if harm results from unexpected use within a

74 Muller (n 32) 22.
75 Solum (n 31) 1274.
76 Mineur (n 11) 22.
77 Vaassen (n 19) 89-90.
78 Mineur (n 11) 22.
79 Soraj Hongladarom and Auriane Van der Vaeren, 'ChatGPT, Postphenomenology, and the Human-Technology-Nature Relations' (2024) 2 Journal of Human-Technology Relations 9. doi:10.59490/jhtr.2024.2.7386.
80 ibid 11-2.
81 Gabriel Hallevy, 'The Criminal Liability of Artificial Intelligence Entities-From Science Fiction to Legal Social Control' (2010) 4(2) Akron Intellectual Property Journal 191.

wide range of possible outcomes, assigning liability becomes more complex.[82] The model's opacity adds to the challenge, as even the developers might not be able to explain why certain content was generated, making it difficult to determine whether the act constituted criminal negligence or an acceptable technical risk.[83]

To address these complexities, the law must go beyond mere adjustments to evidentiary tools. It must construct a new regulatory system that considers the nature of generative artificial intelligence.[84] For instance, clear legal frameworks should be established to govern the development and operation of models like ChatGPT, imposing strict standards for digital safety and governance responsibility, and defining the limits of acceptable risks—or what might be termed the "permissible risk zone".[85]

Imagine a company developing ChatGPT that implements preventive measures such as content filtering and usage management to prevent harmful outputs. These measures would delineate the permissible risk zone, serving as the dividing line between acceptable technical error and liability overreach.[86] Liability would arise only when these boundaries are breached, due to a lack of precautionary measures or willful disregard for potential risks.[87]

Thus, the notions of intent and negligence must be reformulated within a broader framework that accounts for varying degrees of foreseeability and probability, while recognising the unique nature of "actions" generated by a non-conscious mathematical system.[88] The law now faces a fundamental philosophical challenge—not merely redrawing evidentiary tools but reconstructing its foundational perceptions of criminal liability in an era where artificial intelligence has become an independent, complex actor generating actions and consequences that transcend traditional concepts of will and intent.[89]

6 THE LEGAL AND REGULATORY CHALLENGES OF GENERATIVE AI

As generative artificial intelligence, like ChatGPT, grows quickly, new legal issues regarding legal risks and responsibility for generated content surface. Determining criminal liability is challenging due to a regulatory gap left by current frameworks that have not kept up with the rate of innovation. To suggest practical ways to ensure the safe and lawful use of this

82 ibid 194.

83 Yuntao Wang and others, 'A Survey on ChatGPT: AI-Generated Contents, Challenges, and Solutions' (2023) 4 IEEE Open Journal of the Computer Society 297. doi:10.48550/arXiv.2305.18339.

84 Xukang Wang and Ying Cheng Wu, 'Balancing Innovation and Regulation in the Age of Generative Artificial Intelligence' (2024) 14 Journal of Information Policy 397-8. doi:10.5325/jinfopoli.14.2024.0012.

85 ibid 394.

86 Wang and others (n 83) 280.

87 Wang and Wu (n 84) 394-5.

88 Mineur (n 11) 22.

89 Hallevy (n 81) 186-7.

technology, this theme investigates the nature of algorithmic intelligence, evaluates legal risks, and looks at regulatory gaps.

6.1. The Algorithmic Regulatory Gap and Legal Accountability of ChatGPT

Generative AI models, led by ChatGPT, are characterised by a unique cognitive nature, as they operate on complex algorithms that form what is known as the "black box," whose inner workings are difficult to explore or whose outputs are hard to interpret precisely.[90] In the absence of a comprehensive scientific framework clarifying the causal pathways these models follow in making decisions, users face a fundamental challenge: the inability to understand how the model selects certain phrases or information over others, or to interpret the logic underlying the arrangement of elements in the response.[91] For example, when ChatGPT provides a legal recommendation in a criminal case, neither the judge nor the lawyer has access to the detailed reasons that led the model to formulate this recommendation, which obstructs the possibility of accountability if an error or harm occurs.[92] In this sense, this interpretive gap undermines trust in AI outputs and poses fundamental challenges in terms of transparency and legal accountability.[93]

A significant portion of errors produced by models like ChatGPT stems from intertwined technical and legal factors.[94] On the technical level, some errors relate to the design of algorithmic architecture, where certain arrangements in the neural network layers can lead to unintended implicit biases, such as the model discriminating between individuals based on gender or race.[95] At the data level, AI models rely on massive amounts of text drawn from diverse online sources, which may carry cultural or ethical biases or contain inaccurate legal information, thereby transferring these biases into the model's outputs.[96]

Given these multiple challenges, there is an urgent need to develop a strict legal regulatory framework governing the operation of generative AI models and ensures the safety and reliability of their outputs, especially in highly sensitive contexts such as the legal field.[97] The European AI Act represents a pioneering step in this direction,[98] imposing on developers of models like OpenAI a comprehensive risk management obligation, requiring them to ensure training data quality, prepare technical documentation that clarifies decision-

90 Yuan and others (n 20) 5192.
91 Muller (n 32) 3646.
92 ibid 3647.
93 Simmler (n 52) 1146.
94 Amos Azaria, Rina Azoulay and Shulamit Reches, 'ChatGPT is a Remarkable Tool—for Experts' (2024) 6(1) Data Intelligence 241. doi:10.1162/dint_a_00235.
95 Gabbrielle M Johnson, 'Algorithmic Bias: On the Implicit Biases of Social Technology' (2021) 198(10) Synthese 9947. doi:10.1007/s11229-020-02696-y.
96 ibid 9952.
97 Navaneeth (n 7) 62-77.
98 Regulation (EU) 2024/1689 (n 9).

making logic, maintain accurate logs of system outputs, and mandating human oversight of system outputs in high-risk cases.[99] Legal requirements for risk reduction or elimination could make the judiciary less accountable, calling for new models of legal accountability that account for the cognitive differences between humans and machines.[100] This necessitates establishing mechanisms for interpretation and analysis that assist users in consciously and thoughtfully understanding and evaluating the model's results.

6.2. Algorithmic Intelligence and Legitimate Risks

The development of generative artificial intelligence systems, such as the ChatGPT platform, raises novel legal challenges concerning liability for damages resulting from their use.[101] Even if the manufacturing company complies fully with all prescribed technical and regulatory standards, the pressing question remains: Is such compliance sufficient to absolve it of legal liability? Legal and philosophical experience indicates that formal adherence to rules does not necessarily prevent harm, especially when algorithms are granted quasi-autonomous power to make complex decisions, as is the case with these systems operating within unpredictable environments and multifaceted causes.[102] For example, an algorithmic error in interpreting a query or generating inaccurate content may cause serious psychological or social harm that cannot be fully anticipated, thereby fueling debate over whether technical standards alone provide an adequate legal defence.[103]

In legal theory, compliance with rules is recognised as a necessary but insufficient condition for relieving liability, as courts also rely on the "reasonable actor" or "person of similar position and competence" standard.[104] This requires assessing whether the manufacturer took the necessary professional and prudent measures to avoid harm. In complex technical industries, jurisprudence acknowledges the existence of "residual risks" or "acceptable risks"—those that cannot be eliminated except at exorbitant costs or at the expense of technological progress.[105] According to Amy Stein,[106] these risks represent an implicitly accepted zone of risky behaviours, provided they remain within the bounds of control and reasonable precautions, reflecting the reality of technological development and balancing innovation with safety.

99 Kuzma and others (n 8) 1202.
100 Akpobome (n 6) 5051.
101 Johnson (n 95) 9962.
102 Simmler (n 52) 1145.
103 Mike Ananny, 'Seeing like an Algorithmic Error: What Are Algorithmic Mistakes, Why Do They Matter, How Might They Be Public Problem?' (2022) 24(Spec) Yale Journal of Law & Technology 48.
104 ibid 52-3.
105 Azaria, Azoulay and Reches (n 94) 241.
106 Amy L Stein, 'Assuming the Risks of Artificial Intelligence' (2022) 102 Boston University Law Review 983.

On the legislative front, the European Artificial Intelligence Act (AI Act)[107] exemplifies legal evolution that abandons the ideal of eliminating risks and instead adopts a pragmatic philosophy aimed at minimising risks as much as possible while acknowledging their persistent possibility.[108] This law not only imposes strict technical standards but also emphasises algorithmic transparency, continuous system performance monitoring, and the interpretability of decision-making mechanisms.[109] This places a broader duty on manufacturers, extending beyond formal compliance to proactive professional conduct and adaptation to scientific developments.[110] For instance, if damage results from a fault in an algorithm trained on unbalanced data, the manufacturer bears the burden of proving that it took all necessary measures to prevent such errors or mitigate their effects.

The fundamental dilemma lies in delineating the boundary between harm accepted as part of "residual risks" and unjustified failures in design or operation.[111] Accordingly, proposals have emerged to assess liability based on comparing overall system performance against a "safe model" standard; if harmful error rates exceed a certain threshold, the design is considered defective, even if individual errors may be technically justified.[112] However, this approach faces technical challenges related to real-time monitoring, as well as political and ethical considerations involving acceptance of any potential harm to human life.[113] Considering this, there is a pressing need to develop an integrated legal framework that connects technical compliance, professional responsibility, and ethical caution — affirming that justice cannot rest solely on textual adherence but requires a deeper perspective aligned with the complexities and rapid evolution of generative AI.

7 THE FUTURE OF CRIMINAL LIABILITY FOR ALGORITHMIC HARM

Traditional ideas of criminal liability are being called into question by the emergence of generative artificial intelligence, particularly as it permeates business operations. It is becoming more difficult to assign blame for damage brought on by autonomous AI decisions. To predict the future of criminal law amid profound technological change, emerging research aims to reconstruct legal frameworks suited to corporate AI.

107 Regulation (EU) 2024/1689 (n 9).
108 Navaneeth (n 7) 62-77.
109 ibid
110 ibid
111 Stein (n 106) 988.
112 ibid
113 Simmler (n 52) 1145.

7.1. Rebuilding Generative AI's Criminal Liability

Despite notable progress in developing frameworks for criminal liability in the age of artificial intelligence, the legal pathway to holding producers of AI systems, such as ChatGPT, criminally responsible remains complex and ambiguous. Features of these systems—from unpredictability to the opacity of algorithmic decision-making.[114] This creates an epistemic gap between human action and harmful outcome, undermining the establishment of a clear causal link and weakening the logic of blame based on negligence or fault.[115] This phenomenon has been described in legal literature as the "control gap crisis," which criminal law faces in a technological risk society, where traditional models centred on criminal conduct and free will fail to accommodate the emerging complexities of algorithmic actions. For instance, a decision by an autonomous vehicle's algorithm may cause an accident without direct human intervention, raising profound questions about criminal liability.

Traditional criminal liability, which is predicated on intent or negligence, struggles to handle algorithmic decisions. Modern approaches emphasise duty of care violations when AI systems endanger fundamental rights, such as psychological or physical safety, placing a higher priority on preventive liability. Laws such as Article 12 of the UAE consumer protection law No (15) of 2020, which penalise dangerous products but still partially exempts software-only AI, reflect this trend. Thus, liability shifts from fault and intent to adherence to preventive and precautionary obligations.

The particularity of algorithmic harm calls for legislative renewal to keep pace with technological changes by treating AI as a legal entity with tangible effects—even absent physical embodiment. The European AI Act represents a significant step forward, imposing effective, balanced, and deterrent sanctions on violators of AI system regulations.[116] The concept of a "legal warning" model also emerges, criminalising behaviours such as neglecting security measures or failing to update systems, with a combined regime of administrative oversight and criminal penalties to manage risks.[117] Within this framework, authorities may be empowered to issue legal orders mandating additional testing, compulsory updates, or partial system suspension, with violations triggering criminal liability—thereby strengthening societal legal protection.[118]

The nature of algorithmic harm demands establishing a novel legal concept that redefines the relationship between algorithmic acts and legal efficacy, ensuring harm does not escape accountability under the guise of technology and innovation.[119] Accordingly,

114 Vaassen (n 19) 89-90.
115 Simmler (n 52) 1146.
116 Ahkami (n 16) 39.
117 Kuzma and others (n 8) 1201.
118 Cheatham, Javanmardian and Samandari (n 4) 8.
119 Katyal (n 48) 91-2.

enhancing algorithmic transparency by obliging AI producers, such as ChatGPT, to regularly disclose potential risks supports a "democratisation of risk management" approach.[120] This recognises the right of society and legislators to be informed of the potential impacts of technology permeating all aspects of life. Consequently, it becomes essential to combine principles of strict or risk-based liability with advanced preventive legal mechanisms to balance the promotion of innovation with the protection of fundamental legal values such as safety and dignity, thereby achieving legal stability and criminal justice in addressing AI-caused harms.

7.2. Future Studies: Corporate Criminal Law and AI Liability

Traditional ideas of criminal liability are coming under increasing pressure from the emergence of generative artificial intelligence, especially as it becomes increasingly ingrained in business settings.[121] The complexity of determining who is responsible for damage brought about by autonomous AI decisions has led to new research aimed at rebuilding legal frameworks suitable for corporate AI.[122] This changing course aims to anticipate the future of criminal law and guarantee its flexibility in response to the significant technological advancements influencing contemporary responsibility and governance.[123]

Future studies will focus on identifying the parties accountable for damage brought about by generative AI systems in commercial settings.[124] It is expected that questions will arise about the extent to which traditional notions of corporate criminal liability, such as vicarious responsibility or failure to supervise, can address harms caused solely by AI algorithms. It might be necessary to develop new legal frameworks that enable courts to hold companies responsible for algorithmic decisions.[125]

Future research is expected to focus heavily on evaluating businesses' preventive responsibilities when using artificial intelligence. This could mean examining how much corporate governance, risk monitoring, and legal compliance are required of businesses, and how failing to comply could be interpreted as criminal negligence or complicity under corporate criminal law.[126]

Comparative analyses of various legal systems are also likely to become increasingly important, offering insights into how conventional legal doctrines interact with the novel

120 Cheatham, Javanmardian and Samandari (n 4) 8.
121 Mohammad Amin Alkrisheh, 'Criminal Protection of Corporate Websites: An Analytical Study' (2022) 11(3) Journal of Governance and Regulation 148. doi:10.22495/jgrv11i3art12.
122 Osmani (n 58) 59.
123 Kuzma and others (n 8) 1199.
124 Mineur (n 11) 22.
125 Hongladarom and Van der Vaeren (n 79) 11-2.
126 Stavros Kalogiannidis and others, 'The Role of Artificial Intelligence Technology in Predictive Risk Assessment for Business Continuity: A Case Study of Greece' (2024) 12(2) Risks 19. doi:10.3390/risks12020019.

challenges posed by AI-related harms. Such studies may support lawmakers and prosecutors in creating unified regulatory standards and future guidelines to control corporate risks related to artificial intelligence.

The notion of legal capacity for systems like ChatGPT involves treating AI as a potential legal entity with unique rights and responsibilities.[127] At present, global criminal law views AI as a tool for holding humans accountable, even as scholars point out that the existing legal system is not yet fully equipped to address accountability and regulatory questions raised by AI. An industrial or electronic form of legal capacity has been proposed as a potential model for structuring this recognition. Moreover, ChatGPT or AI systems would be punished by operational or technical limitations, such as halting operations or requiring updates, that are intended to prevent harm and ensure compliance, rather than the typical human-centred sanctions.

8 CONCLUSIONS AND RECOMMENDATIONS

The examination of criminal liability for artificial intelligence technologies, especially conversational systems like ChatGPT, reveals profoundly beneficial aspects that transcend the theoretical framework. Legislators are forced to reconsider traditional ideas of criminal liability as well as the *actus reus* (the act) and *mens rea* (intent) components of crime, considering the increasingly intertwined behaviour of humans and machines. It also provides legal professionals with new analytical tools to understand technological risks and assess their obligations in complex digital environments. Therefore, this study contributes to the development of future criminal policy toward a more flexible and equitable system in the age of artificial intelligence, while also improving scholarly discourse.

Fault-based liability is undermined by the difficulty of assigning criminal liability for algorithmic harm, as AI decisions often lack the traditional human intent required for criminal liability. Fundamental ideas in criminal law must be examined in light of the complexity and loss of control over technology. Legal standards that integrate technological tools to enhance preventive oversight and transparency are necessary to ensure accountability for harm caused by AI while balancing innovation and the defence of fundamental rights.

ChatGPT's legal capability could be electronically acknowledged, and operational or technical safeguards could ensure compliance and prevent harm, thereby ensuring accountability. Establishing a direct causal link between algorithmic actions and harmful outcomes is complicated by the opaque, complex nature of AI systems' decision-making processes. Traditional criminal law struggles to effectively address unpredictable, uncontrollable technological risks. There is a growing need to shift from fault-based

127 Yuan and others (n 20).

liability to a preventive liability model that emphasises the failure to implement adequate protective measures. Existing legal frameworks, including consumer protection laws, are inadequate for addressing the unique challenges posed by AI, especially for software that lacks physical integration. Transparency and regulatory oversight are critical to managing AI-related risks, with regular disclosure and mandatory updates serving as key tools to mitigate potential harm.

It is recommended that a legal framework be established that recognises AI's electronic capabilities and outlines operational or technical procedures as accountability mechanisms. Legal frameworks should be modernised to recognise AI as a distinct legal entity, capable of bearing responsibilities regardless of its physical form, with clear duties imposed on developers and users. A preventive liability approach should be adopted, criminalising negligence in failing to ensure safety measures and system updates, with deterrent penalties. AI developers must be required to provide regular transparency reports about potential risks, supporting broader societal and legislative oversight mechanisms. Administrative authorities should be empowered to monitor AI systems, enforce mandatory testing and updates, and impose partial shutdowns, when necessary, with criminal sanctions for non-compliance to ensure effective protection.

REFERENCES

1. Ahkami A, 'AI and the European Union's Approach to Data Protection: The Case of Chat GPT' (Master's thesis, University of Padova 2024)
2. Akpobome O, 'The Impact of Emerging Technologies on Legal Frameworks: A Model for Adaptive Regulation' (2024) 5(7) International Journal of Research Publication and Reviews 5046. doi:10.55248/gengpi.5.1024.3012
3. Alkrisheh MA, 'Criminal Protection of Corporate Websites: An Analytical Study' (2022) 11(3) Journal of Governance and Regulation 148. doi:10.22495/jgrv11i3art12
4. Alneyadi MS and others, 'The Crime of Electronic Blackmail in the Emirati Law' (2022 International Arab Conference on Information Technology (ACIT), Abu Dhabi, UAE, 22-24 November 2022). doi:10.1109/ACIT57182.2022.9994165
5. Al-Uqdah IM, 'Criminal Liability for Artificial Intelligence Application Crimes' (2025) 153 Journal of Law and Jurisprudence 39
6. Ananny M, 'Seeing like an Algorithmic Error: What Are Algorithmic Mistakes, Why Do They Matter, How Might They Be Public Problem?' (2022) 24(Spec) Yale Journal of Law & Technology 342
7. Asil A and Wollmann TG, 'Can Machines Commit Crimes Under US Antitrust Laws?' (2024) 3(1) The University of Chicago Business Law Review 1

8. Atiq EH, 'How Folk Beliefs about Free Will Influence Sentencing: A New Target for the Neuro-Determinist Critics of Criminal Law' (2013) 16(3) New Criminal Law Review 449. doi:10.1525/nclr.2013.16.3.449

9. Azaria A, Azoulay R and Reches S, "ChatGPT is a Remarkable Tool—for Experts' (2024) 6(1) Data Intelligence 240. doi:10.1162/dint_a_00235

10. Beckers S, Chockler H and Halpern J, 'A Causal Analysis of Harm' (Advances in Neural Information Processing Systems 35: 36th Conference on Neural Information Processing Systems (NeurIPS 2022), 28 November - 9 December 2022, New Orleans, Louisiana, USA) 2365

11. Biju PR and Gayathri O, 'Algorithmic Solutions, Subjectivity and Decision Errors: A Study of AI Accountability' (2025) 27(5) Digital Policy, Regulation and Governance 523. doi:10.1108/DPRG-05-2024-0090

12. Blackmore B, 'Looking Beyond Blame and Praise: Analysing Moral Responsibility in the Development and Deployment of AI Systems' (PhD thesis, University of Otago 2023)

13. Carrera S, Mitsilegas V and Stefan M, *Criminal Justice, Fundamental Rights and the Rule of Law in the Digital Age: Report of a CEPS and QMUL Task Force* (CEPS 2021)

14. Cheatham B, Javanmardian K and Samandari H, 'Confronting the Risks of Artificial Intelligence' (2019) 55(2) McKinsey Quarterly 38

15. Chen X, 'Research on the Application of Intelligent ChatGPT in Computer Intelligent Computing System' (2023 IEEE 3rd International Conference on Data Science and Computer Application (ICDSCA), 27-29 October 2023) 985. DOi:10.1109/icdsca59871.2023.10392475

16. Crootof R, '"Cyborg Justice" and the Risk of Technological-Legal Lock-In' (2019) 119(7) Columbia Law Review 233

17. Darlington K, 'Aspects of Intelligent Systems Explanation' (2013) 1(2) Universal Journal of Control and Automation 40. doi:10.13189/ujca.2013.010204

18. Ferrante M, 'Causation in Criminal Responsibility' (2008) 11(3) New Criminal Law Review 470. doi:10.1525/nclr.2008.11.3.470

19. Griffiths O and Thorwart A, 'Effects of Outcome Predictability on Human Learning' (2017) 8 Frontiers in Psychology 511. doi:10.3389/fpsyg.2017.00511

20. Hallevy G, 'The Criminal Liability of Artificial Intelligence Entities-From Science Fiction to Legal Social Control' (2010) 4(2) Akron Intellectual Property Journal 171.

21. Hongladarom S and Van der Vaeren A, 'ChatGPT, Postphenomenology, and the Human-Technology-Nature Relations' (2024) 2 Journal of Human-Technology Relations 1. doi:10.59490/jhtr.2024.2.7386

22. Johnson GM, 'Algorithmic Bias: On the Implicit Biases of Social Technology' (2021) 198(10) Synthese 9941. doi:10.1007/s11229-020-02696-y

23. Kalogiannidis S and others, 'The Role of Artificial Intelligence Technology in Predictive Risk Assessment for Business Continuity: A Case Study of Greece' (2024) 12(2) Risks 19. doi:10.3390/risks12020019

24. Katyal SK, 'Private Accountability in the Age of Artificial Intelligence' (2019) 66(1) UCLA Law Review 54.

25. Kuzma J and others, 'An Integrated Approach to Oversight Assessment for Emerging Technologies' in Marchant GE and Wallach W (eds), *Emerging Technologies: Ethics, Law and Governance* (Routledge 2020) 1197. doi:10.1111/j.1539-6924.2008.01086.x

26. Martins HMG, 'Liability Implications of Artificial Intelligence use in Health: Fault and Risk in Public Sector Healthcare' (Master's thesis, Universidade Catolica Portuguesa 2020)

27. Mineur C, 'Autonomous AI Technology and the Evolution of Legal Personhood in Criminal Law' (Master's thesis, University College Tilburg 2024)

28. Muller S, 'Visual Silence in the Language Portrait: Analyzing young People's Representations of their Linguistic Repertoires' (2022) 25(10) International Journal of Bilingual Education and Bilingualism 3644. doi:10.1080/13670050.2022.2072170

29. Navaneeth M, 'The Need for A Global Regulatory Framework for Artificial Intelligence: Implications of the European Union European Union Artificial Intelligence Act 2024' (Master's thesis, National University of Advanced Legal Studies 2024)

30. Neri R, 'Judging Beyond any Reasonable Doubt: A Logic and Epistemological Rule' (2024) 7 Quaestio Facti: Revista internacional sobre razonamiento probatorio 43. doi:10.33115/udg_bib/qf.i7.23028

31. Osmani N, 'The Complexity of Criminal Liability of AI Systems' (2020) 14(1) Masaryk University Journal of Law and Technology 53. doi:10.5817/mujlt2020-1-3

32. Oviedo-Trespalacios O and others, 'The Risks of Using ChatGPT to Obtain Common Safety-Related Information and Advice' (2023) 167 Safety Science 106244. doi:10.1016/j.ssci.2023.106244

33. Scherer M, 'Artificial Intelligence and Legal Decision-Making: The Wide Open?' (2019) 36(5) Journal of international arbitration 539. doi:10.54648/joia2019028

34. Simmler M, 'Responsibility Gap or Responsibility Shift? The Attribution of Criminal Responsibility in Human-Machine Interaction' (2024) 27(6) Information, Communication & Society 1142. doi:10.1080/1369118X.2023.2239895

35. Sison AJG and others, 'ChatGPT: More than a "Weapon of Mass Deception" Ethical Challenges and Responses from the Human-Centered Artificial Intelligence (HCAI) Perspective' (2024) 40(17) International Journal of Human-Computer Interaction 4853. doi:10.1080/10447318.2023.2225931

36. Solum LB, 'Legal Personhood for Artificial Intelligences' (1992) 70 North Carolina Law Review 1231

37. Stein AL, 'Assuming the Risks of Artificial Intelligence' (2022) 102 Boston University Law Review 979

38. Terzidou K, 'Generative AI for the Legal Profession: Facing the Implications of the Use of ChatGPT Through an Intradisciplinary Approach' (*MediaLaws*, 8 September 2023)

39. Turner J, 'Legal Personality for AI' in Jacob Turner, *Robot Rules: Regulating Artificial Intelligence* (Springer 2018) 173. doi:10.1007/978-3-319-96235-1_5

40. Vaassen B, 'AI, Opacity, and Personal Autonomy' (2022) 35(4) Philosophy & Technology 88. doi:10.1007/s13347-022-00577-5

41. Wang J and others, 'Network Meets ChatGPT: Intent Autonomous Management, Control and Operation' (2023) 8(3) Journal of Communications and Information Networks 239. doi:10.23919/JCIN.2023.10272352

42. Wang X and Wu YC, 'Balancing Innovation and Regulation in the Age of Generative Artificial Intelligence' (2024) 14 Journal of Information Policy 385. doi:10.5325/jinfopoli.14.2024.0012

43. Wang Y and others, 'A Survey on ChatGPT: AI–Generated Contents, Challenges, and Solutions' (2023) 4 IEEE Open Journal of the Computer Society 280. doi:10.48550/arXiv.2305.18339

44. White F, 'Directive 85/374/EEC Concerning Liability for Defective Products: In the Name of Harmonisation, the Internal Market and Consumer Protection' in Giliker P (ed), *Research Handbook on EU Tort Law* (Edward Elgar 2017) 128. doi:10.4337/9781785365720.00013

45. Yan WJ and others, 'Navigating Uncertainties in Machine Learning for Structural Dynamics: A Comprehensive Review of Probabilistic and Non-Probabilistic Approaches in Forward and Inverse Problems' (*arXiv*:2408.08629, 16 August 2024). doi:10.48550/arXiv.2408.08629

46. Youvan DC, 'Reconciling the Probabilistic and Deterministic: Exploring Complexity, Emergence, and Uncertainty in Nature, AI, and Human Cognition' (*Research Gate*, October 2024). doi:10.13140/RG.2.2.16020.10880

47. Yuan Y and others, 'Does ChatGPT Know that it Does Not Know? Evaluating The Black-Box Calibration of Chatgpt' (2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), Torino, Italia, 20-25 May 2024) 5191.

48. Zetzsche DA, Buckley RP and Arner DW, 'The Distributed Liability of Distributed Ledgers: Legal Risks of Blockchain' (2018) 4 University of Illinois Law Review 1361. doi:10.2139/ssrn.3018214

AUTHORS INFORMATION

Raed S A Faqir
PhD, Associate Professor, Criminal Law, College of Law, American University in the Emirates, Dubai, United Arab Emirates.
raed.faqir@aue.ae
Associate Professor in Criminal Law, Faculty of Law, Al-Balqa Applied University, Al-Salt, Jordan,
r.faqir@bau.edu.jo
https://orcid.org/0000-0002-6102-0983
Corresponding author, solely responsible for the manuscript preparing.

Competing interests: No competing interests were disclosed.

Disclaimer: The author declares that his opinion and views expressed in this manuscript are free of any impact of any organizations.

RIGHTS AND PERMISSIONS

EDITORS

ABOUT THIS ARTICLE

Cite this article
Faqir RSA, 'ChatGPT in the Dock: Reflections on the Future of Criminal Liability' (2025) 8(Spec) Access to Justice in Eastern Europe 312-38 <https://doi.org/10.33327/AJEE-18-8.S-o000156>

DOI: https://doi.org/10.33327/AJEE-18-8.S-o000156

Keywords: *generative artificial intelligence, ChatGPT, criminal liability, artificial actor, AI act, legal innovation, algorithmic risk.*

DETAILS FOR PUBLICATION

Date of submission: 19 Aug 2025
Date of acceptance: 30 Oct 2025
Date of Publication: 30 Dec 2025
Whether the manuscript was fast tracked? - Yes
Number of reviewer report submitted in first round: 2 reports
Number of revision rounds: 1 round with conditionally acceptance

Technical tools were used in the editorial process:
Plagiarism checks - Turnitin from iThenticate https://www.turnitin.com/products/ithenticate/
Scholastica for Peer Review https://scholasticahq.com/law-reviews

AI DISCLOSURE STATEMENT

I confirm that no artificial intelligence tools or services were used at any stage of writing, translating, editing, or analyzing content for this manuscript.

Access to Justice in Eastern Europe
ISSN 2663-0575 (Print) ISSN 2663-0583 (Online)
Journal homepage http://ajee-journal.com

АНОТАЦІЯ УКРАЇНСЬКОЮ МОВОЮ

Стаття-думка

CHATGPT НА ЛАВІ ПІДСУДНИХ: РОЗДУМИ ПРО МАЙБУТНЄ ІНСТИТУТУ КРИМІНАЛЬНОЇ ВІДПОВІДАЛЬНОСТІ

Раед С. А. Факір

АНОТАЦІЯ

Вступ. *У цьому дослідженні розглядаються правові проблеми, що виникають через генеративний штучний інтелект. Також було підкреслено обмеження традиційної системи інституту кримінальної відповідальності щодо шкоди, заподіяної результатами роботи ШІ. Дослідження вивчає нові моделі відповідальності для забезпечення підзвітності та захисту прав особи в епоху інтелектуальних машин.*

Генеративний штучний інтелект (ШІ), прикладом якого є ChatGPT, перетворився з простого обчислювального інструменту на когнітивного агента, здатного створювати контент, вирішувати проблеми та ухвалювати рішення. Ця еволюція кидає виклик традиційній системі інституту кримінального права, піднімаючи складні питання про притягнення до відповідальності, у випадках, коли результати, згенеровані ШІ, завдають шкоди або призводять до злочинної поведінки. Дослідження «ChatGPT на лаві підсудних: роздуми про майбутнє інституту кримінальної відповідальності» розглядає ці правові дилеми, спричинені втручанням штучного інтелекту, зосереджуючись на недоліках традиційних концепцій кримінальної відповідальності та вивчаючи потребу в нових правових парадигмах.

Методи. *У дослідженні використано описово-аналітичну та порівняльну методику. Автор здійснив аналіз національного та міжнародного законодавства, правових принципів та сучасної юриспруденції, зосередивши увагу на Європейському законі про штучний інтелект (2024) як моделі. Дослідження вивчає автономні можливості штучного інтелекту, непрозорість алгоритмічного ухвалення рішень і труднощі встановлення причинно-наслідкових зв'язків між діями штучного інтелекту та завданою шкодою. Тематичні дослідження використовуються для вивчення потенційних моделей відповідальності, зокрема превентивної відповідальності та концепції «штучного актора».*

Результати та висновки. *У статті було з'ясовано, що традиційна система інституту кримінальної відповідальності є невідповідною для таких систем штучного інтелекту, як ChatGPT, з огляду на їх часткову автономію та алгоритмічну складність. Також було виявлено потенціал для розширення відповідальності розробників, операторів і користувачів, а також необхідність гнучких правових моделей, які поєднують превентивні, адміністративні та кримінальні заходи. Дослідження підкреслює важливість інтеграції правових інновацій із технологічним наглядом для захисту прав людини, зберігаючи при цьому запобіжну та захисну функції кримінального права.*

Ключові слова: *генеративний штучний інтелект, ChatGPT, кримінальна відповідальність, штучний актор, закон про ШІ, правова інновація, алгоритмічний ризик.*

Access to Justice in Eastern Europe
ISSN 2663-0575 (Print)
ISSN 2663-0583 (Online)
http://ajee-journal.com

CASE STUDY

Case Study

JUDICIAL AI AND THE IRREPARABLE BIAS PROBLEM

Nataliia Mazaraki* and Dmytro Honcharuk

ABSTRACT

Background: *Courts are increasingly experimenting with large language models (LLMs) for tasks such as legal retrieval, drafting support, anonymisation, and triage. Yet the promise of efficiency collides with a structural problem: bias. Human adjudication already reflects cognitive and institutional biases; LLMs trained on past judgments and legal text inherit and sometimes amplify those biases. This article asks a focused question: If AI belongs in courts at all, what is the safe, lawful, and useful lane—especially with respect to bias? The inquiry is situated within fair-trial guarantees and emerging regulatory expectations.*

Methods: *A staged analysis grounded in legal obligations and informed by relevant technical characteristics is employed. First, sources of human and judicial bias are mapped, along with points at which LLMs introduce or magnify bias. Second, hard- and soft-law guardrails relevant to bias control in the justice sector are synthesised. Third, two instructive case studies—COMPAS/Loomis (U.S.) and Ewert v. Canada—are examined to demonstrate how group-level disparities and model opacity can generate due-process risks and to identify remedies transferable to LLM-assisted workflows. Finally, an operational blueprint is derived and applied to identify low-risk, high-yield assistive uses for Ukraine.*

DOI:
https://doi.org/10.33327/AJEE-18-8.S-c000159

Date of submission: 02 Nov 2025
Date of acceptance: 24 Nov 2025
Online First publication: 11 Dec 2025
Last Published: 30 Dec 2025

Results and conclusions: *The analysis shows that fully impartial AI outputs are not attainable in adjudication; bias is ineliminable but can be bounded. For Ukraine, the rational path is to invest first in data curation, secure infrastructure, evaluation capacity, and procurement with audit rights, and to confine AI to retrieval, norm collation, drafting-hygiene checks, and "missed-norms" prompts. The contribution is a governance blueprint that ties specific LLM failure modes to enforceable legal duties and practical safeguards—offering courts a credible, bias-aware lane for AI that improves service while preserving rights.*

1 INTRODUCTION

The legal field is experiencing a cognitive revolution driven by large language models (LLMs). LLMs are changing how courts gather and review case law and write decisions.[1] Research shows that in some jurisdictions, AI technology has begun to provide judgment recommendations in summary proceedings and has, to some extent, improved the efficiency of court document processing (such as China's "Smart Court" project). However, this change may create systemic risks when technological efficiency gains conceal potential cognitive biases.

AI in the judiciary has been explored across multiple contexts: criminal adjudication and sentencing support, bail and parole risk assessment, case-management and triage, judgment drafting and anonymisation, e-discovery, and "smart court" pilots. The research showed measurable efficiency gains in routine tasks but uneven effects on fairness and transparency. Methodologically, the field combines doctrinal analyses of due-process and equality guarantees, case-based audits (e.g., COMPAS/Loomis), technical evaluations of model behaviour, and policy studies of court pilots and procurement.

A consistent theme in this literature is that data-driven tools do not erase human bias; they can mask and reproduce it. O'Neil's early critique framed the problem starkly: data may look objective, but the choices of collection, labelling, modelling and deployment are not[2]—a point echoed in more recent legal analyses.[3] Empirical and technical work converges on three practical bottlenecks for lawful use in courts: data curation (authoritative provenance, representativeness, lawful reuse), document annotation (consistent labelling and redaction at scale), and output validation (citation fidelity, subgroup performance, audit trails).[4] Studies of criminal justice applications report potential gains in speed and consistency, but

1 OECD, *Governing with Artificial Intelligence: The State of Play and Way Forward in Core Government Functions* (OECD Publishing 2025) doi:10.1787/795de142-en.

2 Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (Crown 2016).

3 Qingxia Chen, 'Improving the Trial Efficiency of Criminal Cases with the Assistance of Artificial Intelligence' (2025) 5 Discover Artificial Intelligence 110. doi:10.1007/s44163-025-00353-2.

4 Allison Koenecke and others, 'Tasks and Roles in Legal AI: Data Curation, Annotation, and Verification' (*arXiv preprint*, 2 April 2025) arXiv:2504.01349. doi:10.48550/arXiv.2504.01349.

also underline unresolved issues around privacy, scalability, and algorithmic bias that can translate into group-level disparities if left unmanaged.[5]

Building on this literature, Krištofík offers a court-focused synthesis.[6] He links the ECtHR's subjective and objective tests of judicial impartiality to AI governance, arguing that iterative, life-cycle audits and thorough documentation are the algorithmic counterparts of impartiality guarantees. He reads the CEPEJ Ethical Charter and its assessment tool as the practical route to detect and correct discriminatory patterns. As opacity undermines these safeguards, he calls for outcome monitoring, traceability, and public scrutiny.

Taken together, this record supports a bounded claim: fully impartial AI outputs are not attainable in adjudication, but bias can be constrained when tools are confined to narrow, assistive roles and subjected to evidence-based controls such as representativeness checks, source-grounded retrieval, transparency, logging, subgroup validation, and continuous monitoring. This is the backdrop for this study and motivates the hypothesis tested in the paper.

This paper treats the judiciary as a high-stakes, rights-sensitive environment and asks a focused question: If AI belongs in courts at all, where is the safe, lawful, and useful lane—especially with respect to bias?

To answer this, the study is framed around a few questions that keep bias and fair-trial guarantees at the centre:

1. What makes court AI biased in the first place—in the data, the tools, or the way people use them—and which parts can be reduced versus which are inherent?
2. What do today's European rules and soft-law standards actually allow or forbid for courts, and how do they safeguard reason-giving, equality of arms, and judicial independence?
3. What minimum safeguards are needed?
4. What do practice and cases teach us about the limits of AI in judging and workable remedies?
5. Where is the practical value, especially for Ukraine, and how to keep AI reliable?

5 Lauren E Kois and Preeti Chauhan, 'Criminal Responsibility: Meta-Analysis and Study Space' (2018) 36(3) Behavioral Sciences & the Law 276. doi:10.1002/bsl.2343; Oksana Kaplina and others, 'Application of Artificial Intelligence Systems in criminal Procedure: Key Areas, Basic Legal Principles and Problems of Correlation with Fundamental Human Rights' (2023) 6(3) Access to Justice in Eastern Europe 147. doi:10.33327/AJEE-18-6.3-a000314.

6 Andrej Krištofík, 'Bias in AI (Supported) Decision Making: Old Problems, New Technologies' (2025) 16(1) International Journal for Court Administration 3. doi:10.36745/ijca.598.

2 METHODOLOGY

This study combines legal analysis with technical analysis and targeted case studies, proceeding deliberately from human bias to LLM bias and then to legal guardrails, so the analysis remains anchored in courts' existing fair-trial obligations.

The analysis begins with an examination of human and judicial bias, demonstrating that, to some extent, court judgments themselves reflect these distortions. Bias in LLMs is described as stemming from next-token prediction rather than legal reasoning, with distortions introduced by training data, modelling choices, prompts, context, and interface and automation dynamics.

Subsequently, the relevant hard- and soft-law corpus, together with court-level practice, is elaborated. Empirically, two high-impact case studies (*COMPAS/Loomis and Ewert v. Canada*) are examined as they expose group-level disparities, transparency limits, and judicial remedies that can be generalised to LLM-assisted workflows.

The scope is intentionally normative-operational rather than experimental. Our contribution is a governance blueprint that connects specific LLM failure modes to enforceable legal duties and operational safeguards; empirical validation of any specific tool remains for jurisdiction-level pilots under the outlined controls.

3 BIAS IN THE JUDICIARY: HUMAN, ALGORITHMIC, AND INSTITUTIONAL DIMENSIONS

3.1. A Human Bias

Human judgment is not a neutral measurement device. Under cognitive load, time pressure, or uncertainty, we rely on heuristics[7] that introduce predictable distortions. An early figure (a plea offer, a prosecutor's recommendation, the statutory maximum) anchors subsequent assessments, tugging them toward itself even when arbitrary. Once a working hypothesis forms ("the defendant is likely culpable"), confirmation bias[8] skews attention and memory toward supportive evidence while discounting disconfirming cues, including alternative readings of forensic or eyewitness material. Availability and representativeness replace base-rate reasoning with vivid exemplars and "fitting" crime scripts, inflating perceived likelihoods. Stereotyping means that social cues such as race, age, accent, disability, or socioeconomic status trigger learned associations that subtly shape credibility and

7 Amos Tversky and Daniel Kahneman, 'Judgment under Uncertainty: Heuristics and Biases' (1974) 185(4157) Science 1124. doi:10.1126/science.185.4157.1124.

8 Raymond S Nickerson, 'Confirmation Bias: A Ubiquitous Phenomenon in Many Guises' (1998) 2(2) Review of General Psychology 175. doi:10.1037/1089-2680.2.2.175.

dangerousness judgments, often without conscious endorsement.[9] Knowing outcomes *ex post* (e.g., that harm occurred) fosters hindsight and outcome bias, making events appear more foreseeable and negligence more apparent than they were prospectively. Finally, affective and situational factors: fatigue, time of day, docket pressure, and framing effects (gain vs. loss) systematically shift decisions.[10] Together, these mechanisms operate reliably enough to be mapped and anticipated, which is precisely why courts must treat them as designable risks rather than random chance.

3.2. Judicial Bias: Sources and Safeguards

Impartiality is both a duty and a presumption of judicial office. Judges must approach each case with "an open mind that is free of prejudgment and prejudice."[11] Public confidence in the courts rests on the belief that judges decide with unwavering impartiality. That confidence, in turn, depends on judges' capacity to anticipate and curb the predictable biases of human decision-making.

Scholars commonly distinguish two forms of judicial bias: actual and apprehended (apparent).[12] Actual bias describes a decision-maker who has prejudged the matter or closed their mind, so that relevant, admissible evidence is unlikely to move them—often due to interests, prior conduct or associations, or exposure to extraneous information. As courts are reluctant to probe a judge's inner thinking, the evidential threshold for actual bias is high. By contrast, apprehended bias applies an objective, observer-focused test: Would a fair-minded, informed observer see a real possibility that the judge is not impartial? This approach protects public confidence by focusing on the circumstances that reasonably threaten neutrality rather than proof of subjective prejudice. On this account, bias is not limited to hostility or animus; it includes structural and informational influences capable of undermining open-minded adjudication.

Equality and non-discrimination sharpen the point. Where bias interacts with protected characteristics (sex, race/ethnicity, disability, age, religion, etc.), it risks unjustified disparate treatment (different outcomes for similar facts) or unjustified disparate impact (systematically higher error or burden rates for a group). Even if no one intends discrimination, the law asks whether a differential is relevant and proportionate. As many biases are predictable, they are foreseeable risks that public authorities have a duty to prevent.

9 Patricia G Devine, 'Stereotypes and Prejudice: Their Automatic and Controlled Components' (1989) 56(1) Journal of Personality and Social Psychology 5. doi:10.1037/0022-3514.56.1.5.

10 Shai Danziger, Jonathan Levav and Liora Avnaim-Pesso, 'Extraneous Factors in Judicial Decisions' (2011) 108(17) Proceedings of the National Academy of Sciences 6889. doi:10.1073/pnas.1018033108.

11 Matthew Groves, 'The Rule Against Bias' (2009) 39 Hong Kong Law Journal 486.

12 Gary Edmond and Kristy A Martire, 'Just Cognition: Scientific Research on Bias and Some Implications for Legal Procedure and Decision-Making' (2019) 82(4) Modern Law Review 633. doi:10.1111/1468-2230.12424.

Crucially, judges' biases are often reflected in their judgments and sentencing patterns; when those texts and outcomes are reused as training data, court-facing AI can learn and reproduce the same skews, creating feedback loops. Governance must therefore address both human decision-making and the datasets and models that learn from it.

3.3. Automation Bias

Having outlined human and judicial bias, the study turns to the tools increasingly used around courts: large language models (LLMs). Generative AI is software capable of creating, or generating, various media based on data it has observed in the past and influenced by what people consider pleasing and accurate outputs. More broadly, LLMs are machine learning models trained on large amounts of linguistic data.[13] LLMs do not "know" law; they surface patterns from data. LLMs operate on numbers, not words: input text is tokenised into subword units drawn from a fixed vocabulary, and the model learns statistical relationships among these tokens. Most LLMs are autoregressive, which means that given prior tokens, they predict the next, and a transformer's output is a probability distribution over the vocabulary; one token is then selected by sampling, with a tunable degree of randomness that trades creativity against consistency. As training optimises next-token prediction, not legal reasoning *per se*, capabilities track distributional familiarity: performance is strongest on frequent, well-represented patterns in the training corpus and degrades on genuinely novel tasks.[14] This technical profile underlies both the promise and bias risks of court-facing LLMs.

A court-facing LLM follows a simple life-cycle, and each stage has a characteristic bias risk. It begins with problem framing (what task the tool is meant to help with and which outcomes count as "good"). Choices can pre-tilt results. Next is data building from judgments, legal framework: some parties or case types are barely represented in the record, so the model "learns" mostly from the majority, when anonymisation and redaction is uneven, it may strip out key facts that explain a result but leave clues like postcode, employer, or school and that directs the LLM in a certain way. Finally, case metadata (headnotes, outcome codes) is sometimes incorrect or inconsistently applied. Together, these features teach the system a distorted version of reality and risk perpetuating past inequities. During model adaptation (fine-tuning on legal text), small or skewed in-house datasets may overfit majority writing styles or common fact patterns, sidelining minority languages or rare claims. If the system uses retrieval, ranking that over-weights highly cited or older courts can bias what sources the model sees. In generation, prompt wording and decoding settings favour confident, template-like answers that can normalise stereotypes or default outcomes. Evaluation often misses bias when it reports only average accuracy rather than subgroup

13 Edward Raff, Drew Farris and Stella Biderman, *How Large Language Models Work* (Manning 2025).

14 Dilyan Grigorov, *Introduction to Python and Large Language Models* (Apress 2024). doi:10.1007/979-8-8688-0540-0.

performance or citation fidelity. After deployment, use and interface create an automation bias: time pressure, one-click acceptance, and a lack of disclosure encourage uncritical adoption of AI-generated text. Finally, weak monitoring allows feedback loops, as AI-drafted language re-enters future datasets and hardens the same skew.[15]

4 AI AND LLMS IN COURTS

4.1. Deployment Models and Practices in the Judiciary

Courts are beginning to deploy large language models as assistive tools, not decision-makers. In Portugal, the justice tech agency is piloting AI-based anonymisation of judgments and a "virtual judge assistant" (e.g., STJ's IRIS) for drafting support—uses that require gold-standard tests, human QA, and subgroup error checks. Catalonia has trialled a generative-AI aid for repetitive commercial rulings, framed as assistive only, with disclosure, independent judicial reasons, and "cite-or-abstain" settings to curb hallucinations.[16]

Nationally, Spain permits summarising/drafting support but forbids AI from issuing final decisions. Spain's justice-sector AI policy pivots on two linked duties. First, transparency, impartiality and fairness: systems must be accessible, understandable and auditable. To balance IP with accountability, access to design information should be enabled, along with FAT records (Fairness, Accuracy, Transparency), so that bias can be detected and the interests of justice prevail. Second, prevention of bias and discrimination: algorithms are to undergo periodic evaluations to identify and correct biases arising from training data or model design, with safeguards to protect rights and avoid perpetuating structural injustices. Quality control and auditing scale with risk: where tools might affect the exercise of jurisdiction and judicial independence, oversight lies with the General Council of the Judiciary (CGPJ) under the LOPJ, including "algorithmic surveillance" (collection and analysis of system data/outputs to assess performance, detect bias or errors, and ensure

15 Emily M Bender and others, 'On the Dangers of Stochastic Parrots: Can Language Models be Too Big?' (FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency) 610. doi:10.1145/3442188.3445922; Rishi Bommasani and others, 'On the Opportunities and Risks of Foundation Models' (*arXiv preprint*, 12 July 2022) arXiv:2108.07258. doi:10.48550/arXiv.2108.07258; Xuezhi Wang and others, 'Self-Consistency Improves Chain of Thought Reasoning in Language Models' (*arXiv preprint*, 7 March 2023) arXiv:2203.11171. doi:10.48550/arXiv.2203.11171; Erik Jones and Jacob Steinhardt, 'Capturing Failures of Large Language Models Via Human Cognitive Biases' (NIPS'22: Proceedings of the 36th International Conference on Neural Information Processing Systems) 11785.

16 'Cataluña Prueba Las Sentencias Judiciales Escritas Con Inteligencia Artificial' (*RTVE*, 21 March 2025) <https://www.rtve.es/noticias/20250321/cataluna-sentencias-judiciales-inteligencia-artificial/16502032.shtml> accessed 1 September2025.

transparency and responsibility). The policy explicitly defines model bias: "GenAI tools incorporate any bias from the data sets used to train them… the model output may make systematic errors or favour certain groups, leading to unfair or discriminatory results."[17]

The Netherlands judiciary's AI programme is deliberately incremental: pilots prioritise low-risk, public-facing tools (e.g., website chatbots and internal knowledge search) while standards for court-facing uses are built out. Experiments are sandboxed, use approved corpora, and keep data inside the judiciary infrastructure; logs are retained, and DPIAs/algorithm registers are prepared in anticipation of the EU AI Act duties. Any generative features are framed as assistive (template completion, style harmonisation), with "cite-or-abstain" settings, role-based access, and explicit prohibitions on AI determining outcomes. Staff training covers automation-bias risks and how to read model/system cards; procurement templates require access to documentation and audit rights.[18]

In Singapore, reported pilots in small-claims contexts use gen-AI for summarisation and user guidance, coupled with retrieval from official sources, edit logging, and on-screen disclosures. Outputs are advisory and must be reviewed; acceptance includes friction (checklists/justification prompts), and periodic sampling by registrars checks citation fidelity, neutrality, and subgroup performance. Interfaces are tuned to reduce anchoring (e.g., withholding suggestions until a human outline exists), and prompts/outputs are retained to create an appeal-ready record.[19]

At Brazil's Superior Labour Court (TST), BEM-TE-VI is an AI-assisted triage and case-management tool (not an LLM) used since 2018 to screen incoming labour appeals. It clusters cases by theme, flags timeliness issues, and supports "virtual triage" and routing to analyst teams, speeding cabinet workflows while leaving adjudication to judges.[20]

17 Ministry of the Presidency, Justice and Parliamentary Relations (Spain), *Policy on the Use of Artificial Intelligence in the Administration of Justice* (Preliminary version, MPJRC 2024).

18 Council for the Judiciary (Netherlands), 'Responsible and Innovative: AI for a fair Dutch Judicial System' (*de Rechtspraak*, 2024) <https://www.rechtspraak.nl/Organisatie-en-contact/innovatie-binnen-de-rechtspraak/Paginas/AI-Decree.aspx> accessed 1 September 2025; Ibrahim Jabri, 'The Use of Artificial Intelligence in the Dutch Courtroom' (Master thesis, TU Delft 2022); 'Rotterdam Court Tests Artificial Intelligence as Writing Aid in Criminal Verdicts' (*NL Times*, 30 March 2025) <https://nltimes.nl/2025/03/30/rotterdam-court-tests-artificial-intelligence-writing-aid-criminal-verdicts> accessed 1 September 2025.

19 Supreme Court of Singapore, 'Guide on the Use of Generative Artificial Intelligence Tools by Court Users' (*Singapore Courts*, 2024) <https://www.judiciary.gov.sg/docs/default-source/news-and-resources-docs/guide-on-the-use-of-generative-ai-tools-by-court-users.pdf> accessed 1 September 2025; Maryam Akhlaghi, 'Navigating AI in the Courts: Lessons from Singapore, South Korea, and Australia' (*Laboratoire de Cyberjustice*, 21 July 2025) <https://www.cyberjustice.ca/en/2025/07/21/navigating-ai-in-the-courts-lessons-from-singapore-south-korea-and-australia/> accessed 1 September 2025.

20 Marcos de Moraes Sousa and Thiago Maia Sayão de Moraes, 'Institutionalization of Innovation: The Perception of Actors in the Brazilian Labor Court with Artificial Intelligence' (2025) 27(142) Revista Jurídica da Presidência 293. doi:10.20499/2236-3645.RJP2025v27e142-3215.

Taken together, courts are experimenting with a range of AI tools (from anonymisation and public-facing chatbots to retrieval-grounded drafting aids and, in some systems, triage/classification) deployed as assistive rather than adjudicative technologies. This diversity of use matters as each tool exposes different bias pathways. The next section traces concrete episodes where such biases surfaced in practice, showing how they produced legal and procedural problems—and what those failures teach about safer design and governance.

4.2. Problem Cases: Bias, Opacity, and Due-Process Risks

The reality of bias in court-facing algorithms is now widely documented, and a fast-growing literature dissects dozens of deployments across jurisdictions. To keep this section focused, two illustrative cases that have shaped the debate and judicial practice are highlighted: the U.S. experience with the COMPAS risk tool (and *State v. Loomis*), and Canada's *Ewert v. Canada*. Together, they span different legal systems and decision points (sentencing; security classification/parole), expose distinct bias mechanisms (disparate false-positive rates; lack of subgroup validity for Indigenous offenders), and show the courts' emerging remedies—assistive-only use, independent judicial reasons, transparency about model limits, and requirements for validation and ongoing monitoring. These paired case studies anchor the abstract concern about "AI bias" in concrete adjudicative settings and supply principles drawn on throughout the paper:

1) COMPAS (Correctional Offender Management Profiling for Alternative Sanctions by Northpointe/Equivant) is a proprietary risk-and-needs assessment used by many U.S. jurisdictions at pretrial, sentencing, and supervision stages. It produces scales such as general and violent recidivism risk from questionnaires and criminal-history data; race is not an input, while sex is used for separate norms. The vendor describes COMPAS as empirically developed and validated across jurisdictions, but provides no details on its model design or training data.[21]

In 2016, ProPublica obtained Broward County COMPAS scores and two-year re-arrest outcomes (18,610 people) and reported higher false-positive rates for Black defendants, prompting a national debate.[22] The algorithm's racial bias caused the parole denial rate of African American defendants to be significantly higher than that of white defendants, revealing the falsity of "technological neutrality." The algorithm not only replicates structural discrimination in human society but also may give rise to more hidden

21 EquiVant Supervision, 'Solutions: Risk & Needs Assessments' (*equiVant Supervision*, 2024) <https://equivant-supervision.com/solutions/risk-needs-assessments/> accessed 1 September 2025.

22 Jeff Larson and others, 'How We Analyzed the Compas Recidivism Algorithm' (*ProPublica*, 23 May 2016) <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm> accessed 7 September 2025.

"algorithmic native bias" through its inherent probability model.[23] Northpointe (now Equivant) disputed the analysis, arguing the tool was well-calibrated across groups; subsequent work[24] showed lay predictions can match COMPAS accuracy, sharpening questions about fairness trade-offs and transparency.

In *State v. Loomis*,[25] the Wisconsin Supreme Court allowed the use of COMPAS at sentencing but only with stringent cautions: the score may not be determinative; courts must give independent reasons; and presentence reports must warn of its limitations (proprietary method, group-based design, potential bias). The court acknowledged due-process concerns tied to secrecy but found use permissible under these constraints; the U.S. Supreme Court denied certiorari.

2) In *Ewert v. Canada* (2018 SCC 30),[26] an Indigenous prisoner challenged Correctional Service Canada's use of actuarial risk tools (e.g., PCL-R, VRAG) developed and validated mainly on non-Indigenous populations but applied to inform security classification, programming, and parole. The Supreme Court held CSC breached its statutory duty to ensure information is "as accurate, up to date and complete as possible," as credible evidence of a foreseeable risk of cultural bias existed. CSC had not taken reasonable steps to validate the tools for Indigenous offenders. The Court did not ban actuarial instruments; it required subgroup validation, transparent documentation of the data, populations, and metrics, and ongoing monitoring for disparate error—insisting that any scores be assistive, not determinative, and accompanied by independent reasons. For judicial AI more broadly, Ewert's lesson is straightforward: no black-box deference when fundamental rights are at stake—prove subgroup fairness first, keep humans in control, and continuously audit for bias.

3) Another example comes from the U.S., where two federal district judges were forced to rescind and rewrite rulings after lawyers discovered that the decisions contained non-existent quotations and party descriptions, later traced to the undisclosed use of generative AI by chambers staff. In one case, a New Jersey judge's intern relied on ChatGPT for legal research, producing an order with fabricated case quotations; in another, a Mississippi judge's clerk used the LLM-based tool Perplexity to draft a

23 Li Jialin, 'Exploring Bias Formation Mechanisms in Legal LLMs from a Cognitive Science Perspective' in Xiaofeng Meng and others (eds), *Big Data and Social Computing: BDSC 2025* (Springer 2025) 290. doi:10.1007/978-981-95-0880-8_24.

24 Julia Dressel and Hany Farid, 'The Accuracy, Fairness, and Limits of Predicting Recidivism' (2018) 4(1) Science Advances eaao5580. doi:10.1126/sciadv.aao5580; Tim Brennan and William Dieterich, 'Correctional Offender Management Profiles for Alternative Sanctions (COMPAS)' in Jay P Singh and others (eds), *Handbook of Recidivism Risk/Needs Assessment Tools* (Wiley-Blackwell 2018) 49. doi:10.1002/9781119184256.ch3.

25 *State v Loomis* 2016 WI 68, 371 Wis 2d 235, 881 NW2d 749 (Wis) <https://case-law.vlex.com/vid/state-v-loomis-no-888404547> accessed 6 September 2025.

26 *Ewert v Canada* 2018 SCC 30 [2018] 2 SCR 165 (SCC) <https://decisions.scc-csc.ca/scc-csc/scc-csc/en/item/17133/index.do> accessed 6 September 2025.

temporary restraining order that referenced parties and allegations unrelated to the actual dispute. Only after these errors were called out did the judges acknowledge AI involvement in letters to the Administrative Office of the U.S. Courts, prompting criticism from scholars and a Senate inquiry into the judiciary's AI practices.[27]

5 BIAS CONTROLS IN JUDICIAL AI: HARD- AND SOFT-LAW APPROACHES

Fundamental fair-trial guarantees constrain the use of AI in courts:[28] the right to a fair hearing before an independent and impartial tribunal (ECHR, Art. 6) and the right to an effective remedy and a fair trial (EU Charter, Art. 47). Any AI-assisted workflow must therefore preserve reason-giving, equality of arms, and judicial independence. Where AI tools risk obscuring legal reasoning, skewing access to information, or shifting decision-making authority away from the judge, they jeopardise these guarantees. The next subsection traces how these guarantees are operationalised across key regulatory and soft-law instruments governing AI in the judiciary, showing how fair-trial rights have been translated into concrete requirements and limits for court-facing AI.

Under *the EU Artificial Intelligence Act*,[29] AI systems used by or on behalf of judicial authorities to assist in researching/interpreting facts and law or in applying law to facts are explicitly classified as high-risk (Annexe III, pt. 8(a)). High-risk systems must satisfy the Act's controls: risk management (Art. 9), high-quality data and data-governance (Art. 10), technical documentation (Art. 11), logging/record-keeping (Art. 12), transparency to deployers (Art. 13), human oversight (Art. 14), and accuracy/robustness/cybersecurity (Art. 15), alongside post-market monitoring by providers (Art. 72). Separately, certain practices are banned outright—most notably social scoring that leads to detrimental or unfavourable treatment (Art. 5(1)(c)).

The EU High-Level Expert Group's Trustworthy AI framework[30] tackles bias by hard-wiring it across all seven requirements and operationalising it through the ALTAI self-assessment.[31]

27 Justin Henry, 'Judges Admit to Using AI After Made-Up Rulings Called Out (1)' (*Bloomberg Law*, 23 October 2025) <https://news.bloomberglaw.com/business-and-practice/judges-called-out-for-nonfactual-rulings-admit-to-use-of-ai> accessed 23 October 2025.

28 Iryna Izarova and others, 'Advancing Sustainable Justice Through AI-Based Case-Law Analysis' (2024) 7(1) Access to Justice in Eastern Europe 127. doi:10.33327/AJEE-18-7.1-a000123.

29 Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) <http://data.europa.eu/eli/reg/2024/1689/oj> accessed 1 September 2025.

30 EC High-Level Expert Group on Artificial Intelligence, *Ethics Guidelines for Trustworthy AI* (EU Publications Office 2019). doi:10.2759/346720.

31 EC High-Level Expert Group on Artificial Intelligence, *The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for Self-Assessment* (EU Publications Office 2020). doi:10.2759/002360.

In practice: diversity, non-discrimination and fairness demand representative datasets, scrutiny of sensitive/proxy features, explicit fairness goals, and subgroup error analysis; human agency and oversight is used to counter automation bias with real escalation paths and decision rights for judges; technical robustness and safety requires stress-testing for distribution shift and disparate error rates; privacy and data governance links lawful bases with data-minimisation that avoids proxy discrimination and documents provenance; transparency asks for traceability, explanation and disclosure of limitations so biased patterns can be detected and contested; societal and environmental well-being pushes impact analysis on vulnerable groups; and accountability requires roles, logs, auditability and redress. ALTAI translates this into concrete checks for courts and vendors—e.g., whether training/validation sets are representative, which fairness metrics and trade-offs were chosen (and why), how explanations support bias discovery, what post-deployment monitoring and complaint channels exist, and whether a system should not be deployed if bias cannot be meaningfully mitigated. As procurement and deployment criteria, these items allow judiciaries to require bias audits, subgroup performance reports, mitigation plans, oversight triggers, and ongoing monitoring before any tool is admitted into courtroom workflows.

The Court of Justice of the European Union's AI Strategy treats bias as a first-order risk and organises its entire approach around preventing it without sacrificing judicial independence.[32] It couples strong human control with rigorous risk management that front-loads bias analysis through dataset provenance checks, representativeness testing, and documentation of known limitations. Explainability and traceability are mandated to detect, audit, and correct skewed patterns rather than have them silently propagate into legal reasoning. New systems are introduced only via cautious pilots under close user supervision, which limits automation bias and enables empirical monitoring of subgroup error rates before any wider rollout. Taken together, these elements offer a practical template for national courts: keep AI strictly assistive, design for bias detection and correction from the outset, and make continued use contingent on demonstrable fairness and control.

At the Council of Europe level, the European Ethical Charter on the use of AI in judicial systems (2018) provides five justice-specific guardrails: respect for fundamental rights; non-discrimination with attention to data quality; quality and security; transparency, impartiality and auditability; and the "under user control" principle that keeps AI subordinate to human decision-makers. Crucially, the Charter treats bias as a foreseeable risk: data must be representative and of verified quality; systems should not enable deterministic or black-box outcomes; and design must enable external scrutiny so discriminatory effects can be detected and corrected.[33] Building on this, CEPEJ has shifted

32 Court of Justice of the European Union, *Artificial Intelligence Strategy* (Directorate-General for Information 2023).

33 CEPEJ, *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment* (Council of Europe 2019).

from principles to practice. Its Assessment Tool (2023, updated 2025) operationalises the Charter through concrete checks for discriminatory risk arising from data selection, annotation or training choices, unclear criteria or hidden weightings, and it prescribes mitigation steps and documentation.[34] Where training data or source code cannot be audited, or where bias cannot be ruled out, the guidance counsels against deployment in adjudicative contexts. Complementary 2024 guidance on the online publication of judicial decisions seeks to improve the quality,[35] accessibility, and traceability of the datasets that feed court-facing AI, while the AI Advisory Board's 2025 reporting emphasises regular evaluation and inventory of tools in use.[36]

In parallel, the Council of Europe Framework Convention on Artificial Intelligence (2024) is the first binding, horizontal treaty requiring that AI lifecycle activities comply with human rights, democracy and the rule of law.[37] It expressly recognises that AI can create or aggravate inequalities and therefore imposes duties of risk and impact management, documentation, testing and ongoing monitoring to prevent and mitigate discriminatory outcomes, alongside safeguards for judicial independence and effective remedies for affected persons. Finally, CCJE Opinion No. 26 (2023) sets courtroom-specific guardrails: technology may assist but must not replace adjudication; decision-making remains a human judicial act; design and operation must be non-discriminatory, transparent and intelligible; judges must retain oversight of procurement, design and control; and "judge-analytics" tools that profile or predict individual judges' behaviour are out of bounds.[38] Taken together, these instruments converge on a clear standard: court-facing AI is acceptable only in an assistive, transparent, and auditable role under judicial control, and is not deployed where bias cannot be meaningfully assessed or mitigated.[39]

In EU Member States, standalone AI framework laws are generally unnecessary, as the EU Artificial Intelligence Act already regulates the field and applies directly. What matters, therefore, is how national judiciaries are translating that baseline into practice. Two French

34 CEPEJ, *Assessment Tool for the Operationalisation of the European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment* (CEPEJ(2023)16 final, Council of Europe 2023).

35 CEPEJ, *Guidelines for the Online Publication of Judicial Decisions Aiming at Furthering Legal Knowledge* (CEPEJ(2024)9, Council of Europe 2024).

36 CEPEJ, CEPEJ-GT-Cyberjust and CEPEJ-AIAB, *First Artificial Intelligence Advisory Board (AIAB) Report on the Use of Artificial Intelligence (AI) in the Judiciary* (CEPEJ-AIAB(2024)4Rev5, Council of Europe 2025).

37 Council of Europe, Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law (5 September 2024) [2024] CETS 225.

38 CCJE Opinion No 26 (2023) Moving Forward—The Use of Assistive Technology in the Judiciary (1 December 2023) <https://rm.coe.int/ccje-opinion-no-26-2023-final/1680adade7> accessed 1 September 2025.

39 Tetiana Tsuvina, 'Artificial Intelligence Technologies in the Judiciary: European Standards and Ukrainian Practice' (2025) 139(2) Foreign Trade: Economics, Finance, Law 4. doi:10.31617/3.2025(139)03.

examples stand out. First, the Cour de cassation's joint guidelines for magistrates and lawyers cast generative AI as strictly assistive and treat bias as a primary, foreseeable risk.[40] They require human oversight, independent verification of citations, auditable workflows, user training on bias mechanisms, and the readiness to suspend tools that exhibit discriminatory drift. Second, the Paris Commercial Court's Charter approaches AI "in the spirit of French humanism": it keeps judges responsible for decisions, limits AI to supportive functions, mandates transparency to parties, provides for training and monitored deployment, and centres fairness and non-discrimination.[41]

Ukraine's current AI policy is anchored in cross-sector instruments led by the Ministry of Digital Transformation. The Governmental Concept for AI Development set the baseline for promoting AI R&D, data infrastructure, skills, and ethical principles (transparency, non-discrimination).[42] Still, it did not create sector-specific rules for courts or specify bias controls in adjudication. The Ministerial White Paper on AI Regulation in Ukraine advances regulatory options aligned with EU law, signalling convergence with the EU Artificial Intelligence Act's risk-based model and fundamental rights safeguards.[43] It treats bias as a general risk (fairness, non-discrimination) and anticipates sectoral tailoring, yet it does not provide judiciary-specific operational guidance. The in-progress AI Strategy 2030 (public consultation in 2025) similarly prioritises EU alignment, data/standards, sandboxes, and high-risk governance,[44] but contains no dedicated chapter on the justice sector and no bespoke bias-mitigation regime for court-facing tools.

The research shows that bias in judicial AI is well recognised within the current framework. Hard law (the EU AI Act's high-risk controls and data-protection duties) and soft law (CEPEJ Charter and Assessment Tool, the Council of Europe Framework Convention,

40 Conseil consultatif conjoint de déontologie de la relation magistrats–avocats, 'Intelligence Artificielle Generative: Et vigilance déontologique dans l'exercice professionnel des magistrats et des avocats et de leurs équipes' (*Cour de cassation and Conseil national des barreaux*, 24 October 2025) <https://www.courdecassation.fr/toutes-les-actualites/2025/10/24/magistrats-et-avocats-lere-de-lia-faire-vivre-une-deontologie> accessed 1 September 2025.

41 Vincent Fauchoux, 'From the Paris Edict of 1563 to the AI Act of 2023: How the Paris Commercial Court Regulates Artificial Intelligence in the Spirit of French Humanism' (*DDG Avocats*, 11 September 2025) <https://www.ddg.fr/actualite/from-the-paris-edict-of-1563-to-the-ai-act-of-2023-how-the-paris-commercial-court-regulates-artificial-intelligence-in-the-spirit-of-french-humanism> accessed 1 September 2025..

42 Order of the Cabinet of Ministers of Ukraine No 1556-p 'On the Approval of the Concept for the Development of Artificial Intelligence in Ukraine' (2 December 2020) [in Ukrainian] <https://zakon.rada.gov.ua/go/1556-2020-%D1%80> accessed 1 September 2025.

43 Ministry of Digital Transformation of Ukraine, *White Paper on Artificial Intelligence Regulation in Ukraine: Vision of the Ministry of Digital Transformation of Ukraine* (Version for Consultation, Reforms Delivery Office of the Cabinet of Ministers of Ukraine 2024).

44 Ministry of Digital Transformation of Ukraine, 'The Future of AI in Ukraine Starts Here—Join the Survey' (*Digital State UA*, 27 June 2025) <https://digitalstate.gov.ua/news/govtech/the-future-of-ai-in-ukraine-starts-here-join-the-survey> accessed 1 September 2025.

HLEG "Trustworthy AI"/ALTAI, the CJEU's AI Strategy, and court charters) converge on the same toolkit: rigorous data governance and representativeness checks; meaningful human oversight to counter automation bias; explainability and logging for audit; disclosure to parties; red-lines with a readiness not to deploy where bias cannot be mitigated.

Given these risks, courts are confining AI to assistive roles (research, retrieval, drafting hygiene), not outcome setting or sentencing. Even so, doing assistive AI properly is resource-intensive: it requires upfront investment in secure infrastructure, dataset curation, user training, governance and accountability structures, and then continuous bias monitoring, audits, incident handling, and periodic re-validation over the tool's life cycle.

6 CONCLUSIONS

Bias is not a fringe risk in judicial AI; it is structural. As this paper showed, large language models inherit training-data bias, exhibit inductive/model bias, and are shaped by contextual, retrieval/index, and interface (automation) biases. The case studies (COMPAS/Loomis; Ewert) demonstrate that these mechanisms translate into measurable, group-level disparities and due-process concerns. "Technological neutrality" is therefore a myth: without explicit controls, court-facing AI will reflect and sometimes amplify the unevenness of the legal record from which it learns.

The current hard-and-soft-law framework converges on the same answer: assistive, not determinative use under real human control. The EU AI Act classifies administration-of-justice systems as high-risk and requires data governance, transparency, logging, human oversight, robustness, and post-market monitoring; Council of Europe instruments (CEPEJ Charter/Assessment Tool, Framework Convention) and judicial policies (e.g., CJEU strategy; French court guidance) operationalise these requirements in practice. In short: no black boxes in adjudication; validate on subgroups; disclose limitations; keep an auditable trail; and be prepared not to deploy where bias cannot be meaningfully mitigated.

Done properly, even assistive AI is resource-intensive. It demands secure infrastructure; curated, traceable datasets; gold-standard evaluation (including subgroup performance); procurement with audit rights; user training; and continuous monitoring with a kill-switch. For many systems, especially resource-constrained judiciaries, the rational path is to target low-risk, high-yield uses (retrieval, summarisation, drafting hygiene, anonymisation) and invest first in data quality and basic digitalisation, without which AI will underperform, and budgets will be wasted.

For the Ukrainian realities, in the near term, the most valuable and realistic uses are assistive ones that complement existing national infrastructure rather than replace it. Concretely: (i) case retrieval across Ukrainian jurisprudence, with AI used to surface and cluster relevant Supreme Court lines (not to opine on outcomes)—noting the Supreme Court's increasingly

refined public database; (ii) norm retrieval and collation from the Codes and secondary legislation, recognising that commercial services (e.g., Liga) already deliver strong coverage, so AI should aggregate and cross-check rather than duplicate; (iii) checking judicial drafts for clarity, structure, internal consistency, and citation fidelity; and (iv) "missed-norms" prompts in the Legal Basis section, where a RAG-first tool compares the draft against an authoritative, up-to-date canon of norms used in similar cases and flags omitted but commonly applied provisions for the judge's review. All of these should run inside a logged, court-controlled environment, expose their sources, and leave the final legal reasoning and responsibility solely with the judge.

However, a further constraint is financial. As there are no public disclosures on the cost of judicial-AI deployments anywhere, estimates must be inferred from comparable high-stakes public-sector systems, where initial deployment typically requires €10–20 million, with €1–3 million annually for maintenance, monitoring, and audit.[45] On this basis, the current budget of the Ukrainian judiciary cannot realistically absorb such expenditures, particularly given ongoing reconstruction demands and the still-uncertain marginal benefit that LLM tools would bring to judges' and court staff's workload.

REFERENCES

1. Akhlaghi M, 'Navigating AI in the Courts: Lessons from Singapore, South Korea, and Australia' (*Laboratoire de Cyberjustice*, 21 July 2025) <https://www.cyberjustice.ca/en/2025/07/21/navigating-ai-in-the-courts-lessons-from-singapore-south-korea-and-australia/> accessed 1 September 2025
2. Bender EM and others, 'On the Dangers of Stochastic Parrots: Can Language Models be Too Big?' (FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency) 610. doi:10.1145/3442188.3445922
3. Bommasani R and others, 'On the Opportunities and Risks of Foundation Models' (*arXiv preprint*, 12 July 2022) arXiv:2108.07258. doi:10.48550/arXiv.2108.07258
4. Brennan T and Dieterich Wi, 'Correctional Offender Management Profiles for Alternative Sanctions (COMPAS)' in Singh JP and others (eds), *Handbook of Recidivism Risk/Needs Assessment Tools* (Wiley-Blackwell 2018) 49. doi:10.1002/9781119184256.ch3
5. Chen Q, 'Improving the Trial Efficiency of Criminal Cases with the Assistance of Artificial Intelligence' (2025) 5 Discover Artificial Intelligence 110. doi:10.1007/s44163-02500353-2

45 Vincenzo Piccolo, 'Cost of Implementing AI in Healthcare in 2025' (*Callin.io*, 2025) <https://callin.io/cost-of-implementing-ai-in-healthcare/> accessed 6 September 2025.

6. Danziger S, Levav J and Avnaim-Pesso L, 'Extraneous Factors in Judicial Decisions' (2011) 108(17) Proceedings of the National Academy of Sciences 6889. doi:10.1073/pnas.1018033108

7. Devine PG, 'Stereotypes and Prejudice: Their Automatic and Controlled Components' (1989) 56(1) Journal of Personality and Social Psychology 5. doi:10.1037/0022-3514.56.1.5

8. Dressel J and Farid H, 'The Accuracy, Fairness, and Limits of Predicting Recidivism' (2018) 4(1) Science Advances eaao5580. doi:10.1126/sciadv.aao5580

9. Edmond G and Martire KA, 'Just Cognition: Scientific Research on Bias and Some Implications for Legal Procedure and Decision-Making' (2019) 82(4) Modern Law Review 633. doi:10.1111/1468-2230.12424

10. Fauchoux V, 'From the Paris Edict of 1563 to the AI Act of 2023: How the Paris Commercial Court Regulates Artificial Intelligence in the Spirit of French Humanism' (*DDG Avocats*, 11 September 2025) <https://www.ddg.fr/actualite/from-the-paris-edict-of-1563-to-the-ai-act-of-2023-how-the-paris-commercial-court-regulates-artificial-intelligence-in-the-spirit-of-french-humanism> accessed 1 September 2025.

11. Grigorov D, *Introduction to Python and Large Language Models* (Apress 2024). doi:10.1007/979-8-8688-0540-0

12. Groves M, 'The Rule Against Bias' (2009) 39 Hong Kong Law Journal 485

13. Henry J, 'Judges Admit to Using AI After Made-Up Rulings Called Out (1)' (*Bloomberg Law*, 23 October 2025) <https://news.bloomberglaw.com/business-and-practice/judges-called-out-for-nonfactual-rulings-admit-to-use-of-ai> accessed 23 October 2025

14. Izarova I and others, 'Advancing Sustainable Justice Through AI-Based Case-Law Analysis' (2024) 7(1) Access to Justice in Eastern Europe 127. doi:10.33327/AJEE-18-7.1-a000123

15. Jabri I, 'The Use of Artificial Intelligence in the Dutch Courtroom' (Master thesis, TU Delft 2022)

16. Jialin L, 'Exploring Bias Formation Mechanisms in Legal LLMs from a Cognitive Science Perspective' in Meng X and others (eds), *Big Data and Social Computing: BDSC 2025* (Springer 2025) 290. doi:10.1007/978-981-95-0880-8_24

17. Jones E and Steinhardt J, 'Capturing Failures of Large Language Models Via Human Cognitive Biases' (NIPS'22: Proceedings of the 36th International Conference on Neural Information Processing Systems) 11785

18. Kaplina O and others, 'Application of Artificial Intelligence Systems in criminal Procedure: Key Areas, Basic Legal Principles and Problems of Correlation with Fundamental Human Rights' (2023) 6(3) Access to Justice in Eastern Europe 147. doi:10.33327/AJEE-18-6.3-a000314

19. Koenecke A and others, 'Tasks and Roles in Legal AI: Data Curation, Annotation, and Verification' (*arXiv preprint*, 2 April 2025) arXiv:2504.01349. doi:10.48550/arXiv.2504.01349

20. Kois LE and Chauhan P, 'Criminal Responsibility: Meta-Analysis and Study Space' (2018) 36(3) Behavioral Sciences & the Law 276. doi:10.1002/bsl.2343

21. Krištofík A, 'Bias in AI (Supported) Decision Making: Old Problems, New Technologies' (2025) 16(1) International Journal for Court Administration 3. doi:10.36745/ijca.598

22. Larson J and others, 'How We Analyzed the Compas Recidivism Algorithm' (*ProPublica*, 23 May 2016) <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm> accessed 7 September 2025

23. Moraes Sousa M and Moraes TMS, 'Institutionalization of Innovation: The Perception of Actors in the Brazilian Labor Court with Artificial Intelligence' (2025) 27(142) Revista Jurídica da Presidência 293. doi:10.20499/2236-3645.RJP2025v27e142-3215

24. Nickerson RS, 'Confirmation Bias: A Ubiquitous Phenomenon in Many Guises' (1998) 2(2) Review of General Psychology 175. doi:10.1037/1089-2680.2.2.175

25. O'Neil C, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (Crown 2016)

26. Piccolo V, 'Cost of Implementing AI in Healthcare in 2025' (*Callin.io*, 2025) <https://callin.io/cost-of-implementing-ai-in-healthcare/> accessed 6 September 2025.

27. Raff E, Farris D and Biderman S, *How Large Language Models Work* (Manning 2025)

28. Tsuvina T, 'Artificial Intelligence Technologies in the Judiciary: European Standards and Ukrainian Practice' (2025) 139(2) Foreign Trade: Economics, Finance, Law 4. doi:10.31617/3.2025(139)03

29. Tversky A and Kahneman D, 'Judgment under Uncertainty: Heuristics and Biases' (1974) 185(4157) Science 1124. doi:10.1126/science.185.4157.1124

30. Wang X and others, 'Self-Consistency Improves Chain of Thought Reasoning in Language Models' (*arXiv preprint*, 7 March 2023) arXiv:2203.11171. doi:10.48550/arXiv.2203.11171

AUTHORS INFORMATION

Nataliia Mazaraki*
Dr.Sc. (Law), Prof., International, civil and commercial law, State University of Trade and Economics, Kyiv, Ukraine
Competition law, Max Planck Institut für Innovation und Wettbewerb, Munchen, Germany
n.mazaraki@knute.edu.ua
https://orcid.org/0000-0002-1729-7846
Corresponding author, responsible for conceptualization, research methodology, supervising and writing – original draft.

Dmytro Honcharuk
Master of Law, International, civil and commercial law, State University of Trade and Economics, Kyiv, Ukraine
d.honcharuk@knute.edu.ua
https://orcid.org/0009-0000-8468-6959
Co-author, responsible for data collection and writing – original draft.

Competing interests: No competing interests were disclosed. Any potential conflict of interest must be disclosed by authors.

Disclaimer: The authors declare that their opinion and views expressed in this manuscript are free of any impact of any organizations.

RIGHTS AND PERMISSIONS

EDITORS

Managing Editor – Mag. Yuliia Hartman. **English Editor** – Julie Bold.
Ukrainian language Editor – Mag. Liliia Hartman.

ABOUT THIS ARTICLE

Cite this article
Mazaraki N and Honcharuk D, 'Judicial AI and the Irreparable Bias Problem' (2025) 8(Spec) Access to Justice in Eastern Europe 339-59 <https://doi.org/10.33327/AJEE-18-8.S-c000159>

DOI: https://doi.org/10.33327/AJEE-18-8.S-c000159

Keywords: *artificial intelligence, EU law, judiciary, access to justice, bias.*

DETAILS FOR PUBLICATION

Date of submission: 02 Nov 2025
Date of acceptance: 24 Nov 2025
Online First Publication: 11 Dec 2025
Last Published: 30 Dec 2025
Whether the manuscript was fast tracked? - No
Number of reviewer report submitted in first round: 2 reports (2 external reviewers)
Number of revision rounds: 1 round

Technical tools were used in the editorial process:
Plagiarism checks - Turnitin from iThenticate https://www.turnitin.com/products/ithenticate/
Scholastica for Peer Review https://scholasticahq.com/law-reviews

AI DISCLOSURE STATEMENT

The corresponding author confirms that artificial intelligence tools (such as Grammarly or language editors) were used solely for minor proofreading and readability improvements. All content and intellectual contributions are original and authored by the listed researchers.

АНОТАЦІЯ УКРАЇНСЬКОЮ МОВОЮ

Тематичне дослідження

ШІ В СУДОЧИНСТВІ ТА ПРОБЛЕМА НЕВИПРАВНОЇ УПЕРЕДЖЕНОСТІ

Наталія Мазаракі* та Дмитро Гончарук

АНОТАЦІЯ

Вступ: *Суди дедалі частіше експериментують з великими мовними моделями (LLM) для таких завдань, як пошук правової інформації, допомога в складанні документів, анонімізація та сортування. Однак обіцянка ефективності стикається зі структурною проблемою: упередженістю. Судочинство, яке здійснюється людьми, вже відображає когнітивні та інституційні упередження; LLM, навчені на основі попередніх судових рішень та юридичних текстів, успадковують, а іноді й посилюють ці упередження. У цій статті ставиться конкретне питання: якщо штучний інтелект взагалі має місце в судах, то який шлях є безпечним, законним і корисним, особливо, з*

огляду на упередження? В основі дослідження – гарантії справедливого судочинства та нові регуляторні очікування.

Методи: *У роботі використовується поетапний аналіз, що ґрунтується на правових зобов'язаннях та враховує відповідні технічні характеристики. По-перше, відображаються джерела людської та судової упередженості, а також моменти, у яких LLM вводять або посилюють упередженість. По-друге, синтезуються жорсткі та м'які правові бар'єри, що стосуються контролю упередженості у сфері правосуддя. По-третє, дві показові судові справи — COMPAS/Loomis (США) та Еверт проти Канади — розглядаються, щоб продемонструвати, як диспропорції щодо певних груп людей, та непрозорість моделей можуть створювати ризики для належної правової процедури, та визначити засоби правового захисту, які можна перенести на робочі процеси, що підтримуються LLM. Нарешті, розроблено та застосовано операційний план для визначення низькоризикових та високоефективних способів використання для України.*

Результати та висновки: *Аналіз показує, що повністю неупереджені результати ШІ недосяжні в судовому розгляді; упередженість неможливо усунути, але її можна обмежити. Для України раціональним шляхом є спочатку інвестувати в управління даними, безпечну інфраструктуру, можливості оцінювання та закупівлі послуг ШІ з правами аудиту, а також обмежити використання ШІ пошуком, зіставленням норм, редагуванням та підказками про «пропущені норми». Це дослідження є внеском у план управління, який пов'язує конкретні режими збоїв LLM з юридичними обов'язками, що підлягають виконанню, та практичними гарантіями, пропонуючи судам надійний, орієнтований на вирішення питання упередженості шлях для ШІ, який покращує обслуговування, водночас дотримуючись правових засад.*

Ключові слова: *штучний інтелект, законодавство ЄС, судова система, доступ до правосуддя, упередженість.*

Access to Justice in Eastern Europe
ISSN 2663-0575 (Print)
ISSN 2663-0583 (Online)
http://ajee-journal.com

Case Study

LEGISLATIVE PROTECTION FOR PROPER CRIMINAL JUSTICE PROCEDURES AGAINST PUBLISHING ON SOCIAL MEDIA: A COMPARATIVE ANALYTICAL STUDY

Tayil Shiyab, Hakem Alserhan and Mohammad Alkrisheh*

DOI:
https://doi.org/10.33327/AJEE-18-8.S-c000158

Date of submission: 05 Oct 2025
Date of acceptance: 03 Nov 2025
Online First Publication: 10 Dec 2025
Last Published: 30 Dec 2025

ABSTRACT

Background: *In the digital era, social media platforms have become powerful spaces for public engagement in criminal justice issues, often influencing perceptions of guilt and innocence. Within this context, the Jordanian legal system faces growing challenges in protecting the integrity of judicial processes from the repercussions of premature or prejudicial online publications. This study investigates the extent to which Jordanian law safeguards key procedural guarantees—namely the presumption of innocence, the confidentiality of investigations, and judicial impartiality—against the influence of social media content and public commentary on criminal cases.*

Methods: *The study adopts a doctrinal and comparative legal methodology. It critically analyses the relevant provisions of the Jordanian Penal Code, Cybercrime Law No. 17 of 2023, and Telecommunications Law No. 13 of 1995, alongside related procedural statutes. The research also draws comparative insights from French and Emirati legislation to assess the degree to which Jordanian law aligns with international standards for protecting criminal proceedings from media interference and online bias.*

***Results and conclusions**: Findings reveal that Jordanian legislation does not yet provide sufficiently explicit or comprehensive safeguards to prevent the distortion of justice by public or media pressure. The study highlights significant gaps concerning the protection of investigative confidentiality and the neutrality of judicial authorities. Accordingly, it recommends enacting explicit criminal provisions to prohibit the premature publication of investigative materials, to protect public prosecutors and judges from undue influence, and to increase penalties for unauthorised disclosures. These reforms would strengthen the fairness and independence of the criminal justice system while maintaining a balanced respect for freedom of expression in the digital sphere.*

1 INTRODUCTION

The rise of social media has reshaped the relationship between the public and the criminal justice system. These platforms allow individuals to post, comment on, and circulate crime-related content freely and instantly, often ahead of any judicial determination. While this democratisation of information empowers public discourse, it also introduces significant risks to the administration of justice, particularly concerning the presumption of innocence and the confidentiality of investigations.

In recent years, Jordan has witnessed an increase in the premature dissemination of information regarding criminal cases via social networks. Citizens frequently publish names, photos, and commentary about suspects, which may prejudice judicial outcomes and contribute to what is often referred to as a "trial by media." Such social-media-driven publicity can compromise the fairness of proceedings and erode public confidence in judicial impartiality. These actions challenge the principle of a fair trial and the right to a neutral judiciary.

Despite the growing seriousness of these threats, Jordanian legislation remains fragmented and incomplete. While the Penal Code,[1] the Cybercrime Law No. 17 of 2023,[2] and the Code of Criminal Procedure[3] contain provisions on defamation, confidentiality, and online conduct, none explicitly regulate the publication of ongoing criminal matters in a manner that adequately protects the justice process.

This study arises from a recognised legislative gap: the lack of comprehensive legal safeguards against the undue influence of digital media on criminal trials. The purpose of the research is

1 Jordanian Penal Code No (16) of 1960 [in Arabic] <https://maqam.najah.edu/legislation/33/> accessed 20 September 2025.

2 Jordanian Law No (17) of 2023 'Cybercrime Law' [2023] Official Gazette 5874 [in Arabic] <https://www.assawsana.com/article/602874> accessed 20 September 2025.

3 Jordanian Code of Criminal Procedure of 1961 'Criminal Procedures Law' [1961] Official Gazette 1539 [in Arabic] <https://jordan-lawyer.com/2016/12/22/criminal-law-procedures-jordan-updtaed/> accessed 20 September 2025.

to assess whether Jordanian law sufficiently protects criminal justice procedures from such interference and to propose legal reforms that align with European standards.

The research is structured into four main sections. The first presents the literature review and theoretical framework, identifying scholarly perspectives on the intersection between social media and criminal justice. The second analyses the criminalisation of violations of the presumption of innocence, focusing on how unlawful disclosure or coercive practices undermine this fundamental guarantee. The third examines the criminalisation of acts that influence the proper administration of justice, including prejudicial commentary and media interference during ongoing trials. Finally, the fourth section explores the European context and case law, drawing comparative insights from the jurisprudence of the European Court of Human Rights to support legislative and procedural reform proposals.

2 METHODOLOGY

This study employs a comparative doctrinal methodology supported by qualitative legal analysis. Legal sources were selected purposively according to their direct relevance to the presumption of innocence, judicial impartiality, and the impact of media publications on criminal proceedings.

Primary materials include the Jordanian Penal Code, the Contempt of Courts Act of 1959,[4] and the Jordanian Cybercrime Law No. 17 of 2023, which currently governs online publications and digital expressions related to criminal justice. The study also draws upon the UAE Federal Decree-Law No. 34 of 2021 on Countering Rumours and Cybercrimes,[5] as well as selected judicial decisions of the Abu Dhabi Court of Cassation. These materials were analysed to determine how national laws criminalise prejudicial acts that may distort judicial processes and threaten procedural fairness.

The research further incorporates a European comparative dimension by examining the jurisprudence of the European Court of Human Rights (ECtHR)—notably *Allenet de Ribemont v. France*,[6] *Worm v. Austria*,[7] *Bédat v. Switzerland*,[8] and *Kudeshkina v. Russia*[9]—

4 Jordanian Law No (9) of 1959 'Contempt of Court Law' [in Arabic] <https://cyrilla.org/ar/entity/soilzb8pbvpczuj2ur700ms4i/> accessed 20 September 2025.

5 UAE Federal Decree-Law No 34 of 2021 On Countering Rumors and Cybercrimes [2021] Official Gazette 712 <https://uaelegislation.gov.ae/en/legislations/1526> accessed 20 September 2025.

6 *Allenet de Ribemont v France* App no 15175/89 (ECtHR, 10 February 1995) <https://hudoc.echr.coe.int/eng?i=001-57914> accessed 20 September 2025.

7 *Worm v Austria* App no 22714/93 (ECtHR, 29 August 1997) <https://hudoc.echr.coe.int/eng?i=001-58087> accessed 20 September 2025.

8 *Bédat v Switzerland* App no 56925/08 (ECtHR, 29 March 2016) <https://hudoc.echr.coe.int/fre?i=001-161898> accessed 20 September 2025.

9 *Kudeshkina v Russia* App no 29492/05 (ECtHR, 26 February 2009) <https://hudoc.echr.coe.int/fre?i=001-91501> accessed 20 September 2025.

as well as EU legal instruments, including Article 6 of the European Convention on Human Rights[10] (ECHR), Article 47 of the EU Charter of Fundamental Rights,[11] Directive (EU) 2016/343,[12] and Council of Europe Recommendation Rec (2003)13.[13] These sources were examined through both content-based and normative interpretations to assess how European standards reconcile freedom of expression with fair trial guarantees.

By integrating national and European frameworks, the study employs doctrinal interpretation, comparative synthesis, and critical evaluation to identify legislative gaps, interpretive trends, and reform options that can strengthen the protection of judicial integrity and the presumption of innocence in Jordanian law.

3 LITERATURE REVIEW AND THEORETICAL FRAMEWORK

The rapid development of social media has transformed the dissemination of information related to criminal justice, creating both opportunities and serious challenges for legal systems. Among the primary concerns is the impact of uncontrolled public commentary and digital publication on the presumption of innocence, the secrecy of investigations, and the neutrality of the judiciary. Scholars in comparative criminal law and cyberlaw have increasingly examined these implications, especially in jurisdictions where legislative protections are either insufficient or outdated.

Recent legal scholarship has highlighted how unregulated exposure of criminal cases on social media can violate fair trial guarantees. For example, a 2025 law review note discusses how viral videos and posts shared by non-journalists, often referred to as "TikTok detectives," can undermine the principle of presumption of innocence and pressure judicial authorities.[14] The study argues that the absence of legal safeguards against such exposure jeopardises the impartiality of proceedings and erodes public trust in due process.

Jordan's current legislative framework for addressing digital publication and online commentary during criminal proceedings remains fragmented. Al-Sarayreh analysed the Jordanian Cybercrime Law No. 17 of 2023, concluding that while the law addresses online offences in general, it lacks provisions that specifically shield criminal proceedings from

10 Council of Europe, *European Convention on Human Rights (Convention for the Protection of Human Rights and Fundamental Freedoms, as amended by Protocols)* (ECHR 2013).

11 Charter of Fundamental Rights of the European Union [2012] OJ C 326/391.

12 Directive (EU) 2016/343 of the European Parliament and of the Council of 9 March 2016 On the Strengthening of Certain Aspects of the Presumption of Innocence and of the Right to be Present at the Trial in Criminal Proceedings [2016] OJ L 65/1.

13 Recommendation Rec(2003)13 of the Committee of Ministers to Member States On the Provision of Information Through the Media in Relation to Criminal Proceedings (adopted 10 July 2003) <https://search.coe.int/cm?i=09000016805df617> accessed 20 September 2025.

14 Paige Sanders, 'Protecting the Presumption of Innocence: TikTok Detectives, Misinformation and Private Content Moderation' (2025) 2 Utah Law Review 507.

harmful media exposure.[15] The author emphasised the need for targeted statutory protections that ensure the confidentiality of pretrial proceedings and protect accused individuals from media-driven condemnation.

Khwaileh conducted an empirical study to examine the public perception of the Jordanian Cybercrime Law.[16] His findings suggest that the legal framework fails to align with international standards, such as the European General Data Protection Regulation (GDPR) and the fair trial guarantees enshrined in the International Covenant on Civil and Political Rights (ICCPR). The study advocates for legislative amendments that would strike a balance between privacy protection and transparency without compromising judicial integrity.

Comparative insights from other jurisdictions provide valuable lessons. A 2025 study published in an international law journal examined the phenomenon of "media trials" in India.[17] It concluded that public discourse on criminal matters—when left unchecked—can infringe upon both the freedom of expression and the right to a fair trial. The authors argued for clear statutory limits to prevent public opinion from overwhelming judicial independence.

Similarly, Ligon examined the risks posed by excessive digital transparency in courtroom settings.[18] He observed that the absence of regulations on social media reporting can disrupt the solemnity of criminal trials and shift adjudication from judges to online audiences. The study recommends regulatory frameworks that uphold open justice principles while preserving the integrity of legal proceedings.

The studies reviewed above provide the foundation for this study's theoretical framework, which emphasises the delicate balance between the right to freedom of expression and the need to uphold the integrity of criminal justice procedures. Central to this framework are legal principles such as the presumption of innocence, judicial impartiality, the confidentiality of investigations, and the prohibition against undue influence on judicial processes. The reviewed literature collectively underscores the need for legislative clarity to navigate these competing interests in the digital age.

15 Riyad Mahmoud Al-Sarayreh, 'Jordanian Cybercrime Law No (17) of 2023 between Regulating Social Media Sites and Restricting Freedom of Opinion' (2024) 7(9) Scholars International Journal of Law, Crime and Justice 339. doi:10.36348/sijlcj.2024.v07i09.002.

16 Khaled Mohammad Khwaileh, 'Public Perceptions of Jordan's Cybercrime Law: Protecting Social Media Data and Aligning with International Standards' (2025) 23(1) Pakistan Journal of Life and Social Sciences 2947. doi:10.57239/PJLSS-2025-23.1.00233.

17 PA Anusree and Shampa I Dev, 'The Impact of Social Media Trials on Freedom of Speech and Fair-Trial Rights' (2025) 5(3) International Journal of Applied Law Review.

18 Nicole J Ligon, 'Open Trials in the Social Media Age' (2023) 30(3) Virginia Social Policy Review & the Law 288.

4 INCRIMINATION OF VIOLATION OF THE PRESUMPTION OF INNOCENCE

4.1. Violation of the Presumption of Innocence through Torture and Coerced Confessions

The presumption of innocence is a fundamental principle in criminal justice, ensuring that every accused person is considered innocent until proven guilty by a final, binding court ruling. This principle, enshrined in both national and international legal frameworks, protects the dignity and physical integrity of the accused and prohibits any infringement or harm unless authorised by law.[19]

The meaning of the presumption of innocence is that the accused, no matter how severe his crime is, and no matter how dangerous a criminal he may be, is innocent until legally proven guilty. Accordingly, he shall be treated as innocent, not as a convicted defendant, until the court passes a final, decisive sentence of conviction.[20] The accused does not have to prove his innocence if the Public Prosecution cannot establish evidence. Likewise, this evidence must be firm and decisive, and the trial judge shall be convinced.[21]

On this basis, and through an extrapolation of the Jordanian Penal Code, the researcher has found that the Jordanian legislator has, in many cases, punished the violation of the presumption of innocence.[22] Article 208 of the Penal Code punishes the forcible extraction of statements and information, as it constitutes a clear violation of the presumption of innocence, which states that "the accused is innocent until proven guilty" by a final court ruling issued by the competent court. The same Article stipulated that: (1) whoever inflicts any torture on a person to obtain a confession of a crime or information about it shall be punished by imprisonment from one to three years. (2) For the intent of this Article, torture means any act that results in severe physical or moral pain; or torture which is intentionally inflicted on a person to obtain from him, or another person, information or a confession; or to punish him for an act committed, or suspected of having been committed, by him or

19 Ali Abdul Qader Al-Qahwaji and Sami Abdul Karim Mahmoud, *Principles of Criminology and Penology* (Al-Halabi Legal Publications 2009) [in Arabic]; Maria Stoyanova, 'The Impact of Media Publicity on the Presumption of Innocence' in Lieve Gies (ed), *Trial by Media: Participatory Justice in a Networked World* (Palgrave Macmillan 2025) 269. doi:10.1007/978-3-031-80593-6_11.

20 Fathi Abdul-Ridha Al-Jawari, *The Development of the Iraqi Criminal Justice System* (Legal Research Centre 1986) [in Arabic]; Oksana Khablo and Ivo Svoboda, 'International Standards for the Application of the Presumption of Innocence in Criminal Proceedings' (2024) 29(1) Scientific Journal of the National Academy of Internal Affairs 55. doi:10.56215/naia-herald/1.2024.55; Mikaela Rabinowitz, '"What Will Become of the Innocent?": Pretrial Detention, the Presumption of Innocence, and Punishment Before Trial' (2023) 7(1) UCLA Criminal Justice Law Review 22. doi:10.5070/CJ87162080.

21 Muhammad Muhi al-Din Awad, *Human Rights in Criminal Procedures* (Dar al-Nahda al-Arabiya 1989) [in Arabic].

22 Jordanian Penal Code (n 1).

someone else, to intimidate or coerce this person or others, or, when such pain or suffering is inflicted on the person for any reason, an official, or any person acting in his official capacity, adopts discrimination of any kind, incites it, consents to it, or is silent about it.

The crime of torture is considered a crime against the integrity of the body from a material point of view and its human dignity, as the practice of torture constitutes a serious violation of human rights that are consistent with the principle of the presumption of innocence.

This crime requires a unique criminal intent, represented in the intent to confess, as it is primarily a crime of harm. However, it is an independent crime affecting the proper administration of justice and the integrity of the presumption of innocence.

It is noticeable that the Jordanian legislator did not specify in the aforementioned Article the persons who committed this crime. Hence, the Jordanian legislator equated the officer with the non-officer in the act. The Jordanian legislator should devote a special provision establishing penalties for public servants who employ torture, force, or other means to obtain a confession, as is the case in the UAE legislation.

The Emirati legislator has dedicated a text for the general public and another for public servants.[23] Article 259 of the Federal Penal Code stipulates the punishment of individuals for this act and classifies it as a misdemeanour, as is the case in Jordanian legislation. At the same time, it has devoted another text to public servants, classifying the act as a felony. Article 290 Federal penalties stipulate that "every public servant who uses torture, force or threat, by himself or through the mediation of others, with an accused, witness or expert to induce him to confess to a crime or to give statements or information to conceal a matter, is punished with temporary imprisonment".

According to scholarly analysis, the gravity of this critical crime lies not only in the integrity of the body but also in the proper administration of justice and the uncovering of the truth. In light of this, the Jordanian legislator should include a special provision to punish public servants who commit this act, as per UAE legislation, and to treat it as a felony punishable by a term of imprisonment. This is important given that the consequences of treating it as a felony are entirely different from those arising from treating it as a misdemeanour, such as rehabilitation, prescription, and other processes.

This offence not only threatens the integrity of the body, as previously mentioned, but also jeopardises the progress of the criminal proceedings and the pursuit of the truth. Therefore, a confession lies within the discretionary authority of the trial court. If the court is satisfied with the validity of the confession and convinced of its truth, it may rely upon it. In contrast, if the court is not convinced of the confession, and it is established that it was taken under duress, it must exclude it from the evidence list, as such evidence cannot be relied upon. In such cases, the court must continue to hear the evidence to form its conscience conviction.

23 UAE Federal Law No 3 of 1987 'Promulgating the Penal Code' [1987] Official Gazette 182, and amendments.

Therefore, in its ruling on this matter, the Abu Dhabi Court of Cassation ruled that the plea to invalidate a confession on the grounds of coercion, together with a request to refer the case to a forensic doctor to prove material coercion, constitutes a substantial defence, with a corresponding response being required. Violating this constitutes a deficiency and a breach of the right of defence.[24]

Among the crimes that affect the presumption of innocence and the proper administration of justice is the disclosure of the secrets of a criminal investigation. Criminal legislation has ensured that criminal investigations are confidential, and their contents must not be disclosed. The rationale for confidentiality lies in protecting the interests of the investigation, on the one hand, and preserving the innocence and reputation of the accused, on the other hand, by avoiding exposure of the accused person and their reputation, as well as influencing witnesses.[25]

It is impermissible to interfere with an ongoing investigation by disseminating its details to the public, commenting on them, taking notes, or using them as informational material for individuals to share on social networks. This aligns with recent comparative analyses of digital offences in Emirati law, which recognise that misuse of digital applications and online platforms can constitute criminal interference with justice when such tools are used to disseminate prejudicial or confidential information.[26] Individuals who repeatedly and widely publish details of an investigation via such networks and comment on them pose a grave danger to the freedom and reputation of the individuals whose destinies have led them to the courts as a result of error or injustice. Ultimately, it is only after judicial scrutiny that the truth becomes clear to the judges.[27]

This misuse of social media platforms parallels broader cyber-offences addressed in comparative legal research, which demonstrate how digital communication tools may be exploited to exert pressure, distort reputations, or influence legal proceedings—issues that equally endanger the presumption of innocence and the impartial administration of justice.[28]

24 Appeal No 95 of 2012 (Penal) [2012] Abu Dhabi Court of Cassation.

25 Fair Trials, *Presumption of Innocence and Right to be Present at Trial Directive: Presumption of Innocence, Right to Remain Silent and Not to Incriminate Oneself, Right to Be Present at the Trial* (Fair Trials 2020) <https://www.fairtrials.org/the-right-to-a-fair-trial/the-presumption-of-innocence/> accessed 20 September 2025; Rabinowitz (n 20).

26 Ali Sultan Ali and Raed SA Faqir, 'Criminal Protection of Digital Applications in the UAE Legislation: A Comparative Study' (2024) 16(1) Pakistan Journal of Criminology 489. doi:10.62271/pjc.16.1.489.504.

27 Roger Merle and Andre Vitu, *Traité de droit criminel: Procédure pénale* (5th edn, Dalloz 2000) vol 2.

28 Mohammed Salem Alneyadi and others, 'The Crime of Electronic Blackmail in the Emirati law' (2022 International Arab Conference on Information Technology (ACIT), Abu Dhabi, UAE, 22-24 November 2022). doi:10.1109/ACIT57182.2022.9994165.

The Jordanian legislator has stipulated that the disclosure of investigation secrets is prohibited to protect the interests of the individual and the investigation, as well as to safeguard public confidence in certain professions. This prohibition reflects a broader social interest in preserving secrets and protecting public opinion from undue influence due to the publication of news about the crime and the investigation. To ensure the effectiveness of this protection, liability is not limited to criminal liability but also extends to disciplinary responsibility.[29]

Torture directly undermines the presumption of innocence by transforming the accused from a rights-bearing individual into a source of forced evidence, thereby reversing the burden of proof and destroying judicial impartiality.

In conclusion, coerced confessions represent a dual infringement against human dignity and against the procedural guarantee that every person must be presumed innocent until proven guilty. Such acts distort justice at its very foundation and justify the need for strict statutory provisions to deter them.

4.2. Violation of the Presumption of Innocence through Disclosure of Investigative Secrets

The unauthorised disclosure of investigative information undermines the presumption of innocence.[30] Because it publicly associates the suspect with criminal behaviour before judicial verification. Such premature exposure creates bias, shapes public opinion, and may pressure judicial actors, thereby threatening the neutrality of the criminal process.

Hence, both torture-induced confessions and the disclosure of investigative secrets constitute distinct yet interconnected forms of violating the presumption of innocence. Each disrupts the balance between the rights of the accused and the public's right to information. These violations pave the way for broader media-related offences, which will be examined in the following section on Incrimination of Influencing the Proper Administration of Justice.[31]

29 Abdulaziz Alhassan and others, 'Substantive Criminal Protection for the Right to Image in the Digital Era under UAE and French Legislations' (2024) 7(2) Access to Justice in Eastern Europe 325. doi:10.33327/AJEE-18-7.2-a000216.

30 *Allenet de Ribemont v France* (n 6).

31 Ferry de Jong and Leonie van Lent, 'The Presumption of Innocence as a Counterfactual Principle' (2016) 12(1) Utrecht Law Review 32. doi:10.18352/ulr.324.

5 INCRIMINATION OF INFLUENCING THE PROPER ADMINISTRATION OF JUSTICE

Achieving justice requires that every citizen has the right to resort to a fair and impartial judge who is only affected by right and justice.[32] The supreme objective of the judiciary is to establish justice, which reassures the soul and secures the individual's dignity, property, and life.[33] With such reassurance, people feel secure enough to build and contribute to society. The judge symbolises judicial justice, and their pursuit makes truth attainable. The judiciary holds a unique status among public authorities due to its role in maintaining social stability, settling disputes, protecting freedoms, and upholding justice. For this reason, it must function independently and without external influence. No authority governs the judiciary except for the law.[34]

Based on this, to ensure the judiciary can perform its role free from undue influence, legislators have criminalised actions that obstruct the proper functioning of the justice system. Such offences are serious because they undermine the foundations of judicial integrity and public confidence. The increasing integration of digital and telecommunication technologies into criminal proceedings has further complicated this issue, requiring legislators to establish clear procedural safeguards that preserve judicial impartiality and protect data integrity in the digital age.[35] In this context, modern scholarship on digital criminal investigations highlights that the rise of artificial intelligence and automated data analysis introduces both opportunities and legal vulnerabilities in the pursuit of justice. Ensuring the reliability, transparency, and procedural legitimacy of AI-assisted investigations is now a crucial dimension of judicial protection.[36] The Jordanian Penal Code outlines several such offences:

- Article 196(3) punishes insults or threats directed at a judge during a session with imprisonment from six months to two years.
- Article 223 criminalises submitting petitions or verbal arguments to influence a judge's decision.

32 Adnan Ajil Obaid, The Impact of the Independence of the Judiciary from Conflict in the State of Law: A Comparative Constitutional Study between Arab and International Regulations (Arab Centre for Publishing and Distribution 2018) [in Arabic].

33 Ahmed Ali Alnuaimi and Mohammad Amin Alkrisheh, 'Advancing Criminal Justice through Mediation: Analysing the Integration of Mediation in Emirati Criminal Legislation' (2024) 11 Humanities and Social Sciences Communications 927. doi:10.1057/s41599-024-03458-8.

34 Obaid Muhayer Alshamsi and Mohammad Amin Alkrisheh, 'Criminal Settlement Provisions in Emirati Legislation' (2024) 16(3) Pakistan Journal of Criminology 1073.

35 Mohammad Alkrisheh, 'Guarantees to the Use of Telecommunication Technology in Criminal Proceedings in the United Arab Emirates Legislation' (2020) 10(4) The Lawyer Quarterly 509.

36 Raed S A Faqir, 'Digital Criminal Investigations in the Era of Artificial Intelligence: A Comprehensive Overview' (2023) 17(2) International Journal of Cyber Criminology 77.

- Article 224 punishes publishing news or criticism that may impact a judge or witness or discourage testimony.
- The Contempt of Courts Act (1959), Article 11, criminalises publications that influence judges, prosecution, witnesses, or public opinion regarding pending cases.
- Telecommunications Law No. 13 of 1995, Article 75(a) penalises threats, insults, and the dissemination of false news via communication channels.

Article 196(3) requires that the insulting or threatening conduct be committed in court, and that it addresses the judge directly during proceedings. While this protects judges, it excludes prosecutors who also fulfil a judicial role. To ensure the complete protection of the judicial process, it is recommended that these protections be extended to members of the Public Prosecution, as they are also integral to upholding justice.

Furthermore, Articles 223 and 224 should be amended to include prosecutors. This ensures that any act aimed at influencing a legal proceeding, whether before a judge or a prosecutor, is equally punishable. For instance:

- Article 223 should read: "Everyone who submits a petition to a member of the Public Prosecution or a judge... shall be punished..."
- Article 224 should be amended to read: "Whoever publishes news, information, or criticism that may affect any judge, prosecutor, or witness..."

Publishing that interferes with justice is not limited to negative remarks. Excessive praise can also bias judicial conduct. Judicial influence via social media, whether through rumour, exaggerated support, or criticism, poses a serious threat to judicial neutrality and public trust.

Foreign legal systems recognise this danger. Courts in the UK, Canada, and the US have ruled that media content predicting or praising court outcomes constitutes contempt of court. For example:

- In the UK, commentary on pending cases may be seen as interference.[37]
- In Canada, articles assuming guilt or predicting trial outcomes were ruled contemptuous.
- In the US, a newspaper article implying the accused was certainly guilty and urging public acceptance of this was deemed a direct insult to the judiciary.[38]

37 UK Contempt of Court Act 1981, c 49 <https://www.legislation.gov.uk/ukpga/1981/49> accessed 20 September 2025; Eric Barendt (ed), *Media Freedom and Contempt of Court* (Routledge 2009). doi:10.4324/9781315091297.

38 ABA, 'Judicial Ethics, Impartiality, and the Media' (American Bar Association, 12 August 2019) <https://www.americanbar.org/groups/judicial/resources/judges-journal/archive/judicial-ethics-impartiality-media/> accessed 20 September 2025.

Even praise may influence judges, foster arrogance, or affect impartiality. Judges must avoid both excessive severity and undue leniency. Justice demands clarity, fairness, and emotional neutrality. Hatred or bias—even unconscious—can compromise verdicts.[39]

Finally, individuals must refrain from publishing or commenting on material that may influence judicial proceedings. Reporting should be limited to factual crime news and public trial outcomes. Confidential investigations and private hearings must remain undisclosed to safeguard justice and individual rights. This balance ensures that the public stays informed without violating privacy, dignity, or the presumption of innocence.

6 EUROPEAN CONTEXT AND CASE LAW

Having analysed the Jordanian legal framework governing the publication of criminal justice information on social media, it is essential to situate these findings within the broader European context. European human rights law, particularly the jurisprudence of the European Court of Human Rights (ECtHR), has developed a rich body of principles concerning the balance between freedom of expression and the right to a fair trial. A comparative analysis of European standards provides valuable insights to inform Jordanian legal reform, particularly regarding the presumption of innocence, judicial impartiality, and the confidentiality of criminal investigations.

6.1. European Legal Standards on Fair Trial and Media Freedom

At the European level, the right to a fair trial is primarily enshrined in Article 6 of the European Convention on Human Rights (ECHR), which guarantees the presumption of innocence and the right to a fair trial before an independent and impartial tribunal.[40] Similarly, Article 47 of the Charter of Fundamental Rights of the European Union (EU Charter) ensures the right to an effective remedy and a fair hearing before an independent court.[41] Directive (EU) 2016/343 on the presumption of innocence and the right to be present at trial sets minimum rules to prevent public officials or authorities from presenting suspects as guilty before conviction.[42] Furthermore, the Council of Europe's Recommendation Rec(2003)13 on the provision of information through the media in relation to criminal proceedings outlines the delicate balance between the public's right to

39 Mohammad Amin Alkrisheh and Fatiha Mohammed Gourari, 'Criminal Liability for Paid Disinformation in the Digital World: A Comparative Study Between UAE Law and The European Digital Services Act (DSA)' (2025) 8(2) Access to Justice in Eastern Europe 341. doi:10.33327/AJEE-18-8.2-r000110; Stoyanova (n 19).

40 Council of Europe (n 10) art 6(1)-(2).

41 Charter of Fundamental Rights (n 11) art 47.

42 Directive (EU) 2016/343 (n 12).

information and the need to protect ongoing judicial processes.[43] Together, these instruments articulate a framework in which restrictions on speech are permissible only when proportionate, necessary, and directed at safeguarding the fairness of the proceedings.

6.2. ECtHR Jurisprudence on Balancing Media Freedom and Fair Trial Rights

The ECtHR has consistently recognised the potential of media reporting and official statements to prejudice the fairness of trials. In *Allenet de Ribemont v. France*, the Court found a violation of Article 6(2) ECHR after senior police officials publicly referred to the applicant as one of the instigators of a murder prior to any conviction,[44] thereby undermining the presumption of innocence. Similarly, in *Daktaras v. Lithuania* and *Butkevičius v. Lithuania*, the Court reaffirmed that public authorities must refrain from making statements that could be interpreted as declarations of guilt while proceedings are pending.[45]

Conversely, in *Worm v. Austria*, the Court upheld the conviction of a journalist who had commented on an ongoing criminal case, holding that such expression could endanger the impartiality of the judiciary and the presumption of innocence of the accused.[46] This reasoning was later refined in *Bédat v. Switzerland*, where the Grand Chamber held that imposing a criminal fine on a journalist for publishing confidential investigative material did not breach Article 10 of the ECHR, as it pursued the legitimate aim of preserving the authority of the judiciary and protecting the private life of the accused.[47]

Finally, in *Kudeshkina v. Russia*, the Court underlined that judicial independence requires safeguarding judges from disciplinary or political pressure, while also ensuring that public communication by judges does not compromise public confidence in the judiciary.[48]

6.3. Comparative Perspective: Jordan and Post-Social Legal Transitions

The jurisprudence of the ECtHR is particularly relevant to post-social and transitional legal systems that seek to balance freedom of expression with the integrity of judicial institutions. In Eastern European contexts such as Lithuania, Russia, and Poland, the Court's case law has served as a guiding framework for achieving a balance between open

43 Recommendation Rec(2003)13 (n 13).
44 *Allenet de Ribemont v France* (n 6).
45 *Daktaras v Lithuania* App no 42095/98 (ECtHR, 10 October 2000) ECHR <https://hudoc.echr.coe.int/eng?i=001-58855> accessed 20 September 2025; *Butkevičius v Lithuania* App no 48297/99 (ECtHR, 26 March 2002) <https://hudoc.echr.coe.int/eng?i=001-60344> accessed 20 September 2025.
46 *Worm v Austria* (n 7).
47 *Bédat v Switzerland* (n 8).
48 *Kudeshkina v Russia* (n 9).

justice and impartial adjudication. Jordan's experience—where public commentary and online discourse can influence perceptions of guilt—reflects similar transitional challenges. As in these European jurisdictions, there is a need for codified prohibitions on prejudicial statements by public authorities, regulation of investigative secrecy, and proportionate penalties for unauthorised disclosures. Aligning Jordanian provisions with European standards would enhance both judicial independence and public trust in criminal proceedings.[49]

Drawing on European jurisprudence, the study concludes that Jordanian legislation would benefit from adopting explicit rules on neutral communication by public authorities, proportional limits on media coverage, and codified investigative secrecy aligned with the EU Directive (2016/343) and the ECtHR standards. Such reforms would strengthen judicial independence and public confidence in the justice system.

7 CONCLUSIONS

The findings of this research confirm the serious risks posed by public exposure and social media engagement in ongoing investigations and secret judicial proceedings. Publishing investigative materials, images of defendants or witnesses, or commentary on ongoing trials interferes with the judiciary's independence and impairs the presumption of innocence. Recent comparative legal scholarship has highlighted that restorative justice mechanisms, such as criminal reconciliation, contribute to reinforcing judicial integrity and public trust by promoting fairness, transparency, and community-based dispute resolution.

Such acts represent a tangible threat to the integrity and confidentiality of the judicial process and must be subject to stringent regulation.

The rapid spread of information on digital and social networking platforms amplifies these risks exponentially. Unlike traditional media, social media enables instantaneous and widespread dissemination of unverified information, often accompanied by emotionally charged public commentary. This dynamic environment blurs the boundary between public opinion and judicial reasoning, creating parallel "digital trials" that can prejudice judges, prosecutors, and even witnesses before a verdict is reached.

49 European Commission, *Report from the Commission to the European Parliament and the Council: On the Implementation of Directive (EU) 2016/343 of the European Parliament and of the Council of 9 March 2016 on the Strengthening of Certain Aspects of the Presumption of Innocence and of the Right to be Present at the Trial in Criminal Proceedings* (COM/2021/144 final, 31 March 2021) <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2021:144:FIN> accessed 20 September 2025; European Union Agency for Fundamental Rights, *Presumption of Innocence and Related Rights: Professional Perspectives* (FRA, 2021) <https://fra.europa.eu/en/publication/2021/presumption-of-innocence> accessed 20 September 2025.

Media publication—whether critical or overly supportive—can create undue pressure on judges and the public prosecution. These influences can distort judicial conscience and undermine public confidence in legal outcomes. Although international legal systems, such as the French model, emphasise the judge's inner conviction as a safeguard against external influence, comparative analysis reveals gaps in Jordanian legislation on this issue.

One of the key findings of this study is the disparity in legal protection offered to judges and members of the public prosecution in Jordan. Article 196(3) of the Penal Code provides explicit protection for judges but excludes prosecutors, which represents a negative legislative distinction. Similarly, Articles 223 and 224 fail to impose criminal liability on those who attempt to influence prosecutors, as they do on judges. This imbalance undermines the prosecution's institutional integrity, a key pillar in the pursuit of justice.

Additionally, the study finds that the Jordanian Penal Code, particularly Article 225, imposes insufficient penalties for the publication of confidential judicial documents. The penalties are minimal and do not deter the publication of information that should remain confidential. This, in turn, facilitates harmful practices that compromise due process and the dignity of persons involved in criminal proceedings.

The absence of criminalisation of online publication of sensitive legal materials—such as investigation records, names or photos of suspects, and details of closed-door hearings—under the Jordanian Cybercrime Law is another critical legislative gap. Compared to the Emirati approach, which draws a clear distinction between public and official actors in defining and punishing such offences, Jordanian law appears underdeveloped and outdated.

In light of these deficiencies, the study concludes that Jordanian criminal law requires a systematic revision to align with contemporary challenges posed by media and digital technologies. The law must ensure that both judges and prosecutors are equally protected from attempts to influence their impartiality. This approach aligns with broader international legal perspectives that emphasise transparency and institutional integrity as essential safeguards of justice.[50] Furthermore, any act that interferes with the proper administration of justice—especially those involving the publication of confidential information—must be met with practical legal consequences.

Ultimately, safeguarding the presumption of innocence in the digital era requires more than procedural amendments—it necessitates a cultural and regulatory shift that acknowledges the pervasive influence of social media on the justice system. Establishing clear legislative boundaries for online expression in criminal matters, while respecting freedom of speech, will strengthen public trust in the judiciary and ensure that the principles of fairness and impartiality remain intact in the digital age.

50 Jamal Barafi and others, 'Anti-Corruption Mechanisms: A Study in the Light of International Law and National Regulations' (2022) 11(4) Journal of Governance and Regulation 224. doi:10.22495/jgrv11i4siart3.

Only through these reforms can the Jordanian legal system fulfil its constitutional mandate to safeguard justice, dignity, and individual freedoms against the influence of public discourse and technological abuse.

REFERENCES

1. Alhassan A and others, 'Substantive Criminal Protection for the Right to Image in the Digital Era under UAE and French Legislations' (2024) 7(2) Access to Justice in Eastern Europe 325. doi:10.33327/AJEE-18-7.2-a000216
2. Ali AS and Faqir RSA, 'Criminal Protection of Digital Applications in the UAE Legislation: A Comparative Study' (2024) 16(1) Pakistan Journal of Criminology 489. doi:10.62271/pjc.16.1.489.504
3. Al-Jawari FAR, *The Development of the Iraqi Criminal Justice System* (Legal Research Centre 1986) [in Arabic]
4. Alkrisheh M, 'Guarantees to the Use of Telecommunication Technology in Criminal Proceedings in the United Arab Emirates Legislation' (2020) 10(4) The Lawyer Quarterly 509
5. Alkrisheh MA and Gourari FM, 'Criminal Liability for Paid Disinformation in the Digital World: A Comparative Study Between UAE Law and The European Digital Services Act (DSA)' (2025) 8(2) Access to Justice in Eastern Europe 341. doi:10.33327/AJEE-18-8.2-r000110
6. Alneyadi MS and others, 'The Crime of Electronic Blackmail in the Emirati law' (2022 International Arab Conference on Information Technology (ACIT), Abu Dhabi, UAE, 22-24 November 2022). doi:10.1109/ACIT57182.2022.9994165
7. Alnuaimi AA and Alkrisheh MA, 'Advancing Criminal Justice through Mediation: Analysing the Integration of Mediation in Emirati Criminal Legislation' (2024) 11 Humanities and Social Sciences Communications 927. doi:10.1057/s41599-024-03458-8
8. Al-Qahwaji AAQ and Mahmoud SAK, *Principles of Criminology and Penology* (Al-Halabi Legal Publications 2009) [in Arabic]
9. Al-Sarayreh RM, 'Jordanian Cybercrime Law No (17) of 2023 between Regulating Social Media Sites and Restricting Freedom of Opinion' (2024) 7(9) Scholars International Journal of Law, Crime and Justice 339. doi:10.36348/sijlcj.2024.v07i09.002
10. Alshamsi OM and Alkrisheh MA, 'Criminal Settlement Provisions in Emirati Legislation' (2024) 16(3) Pakistan Journal of Criminology 1073
11. Anusree PA and Dev SI, 'The Impact of Social Media Trials on Freedom of Speech and Fair-Trial Rights' (2025) 5(3) International Journal of Applied Law Review
12. Awad MM, *Human Rights in Criminal Procedures* (Dar al-Nahda al-Arabiya 1989) [in Arabic]

13. Barafi J and others, 'Anti-Corruption Mechanisms: A Study in the Light of International Law and National Regulations' (2022) 11(4) Journal of Governance and Regulation 224. doi:10.22495/jgrv11i4siart3

14. Barendt E (ed), *Media Freedom and Contempt of Court* (Routledge 2009). doi:10.4324/9781315091297

15. de Jong F and van Lent L, 'The Presumption of Innocence as a Counterfactual Principle' (2016) 12(1) Utrecht Law Review 32. doi:10.18352/ulr.324

16. Faqir RSA, 'Digital Criminal Investigations in the Era of Artificial Intelligence: A Comprehensive Overview' (2023) 17(2) International Journal of Cyber Criminology 77

17. Khablo O and Svoboda I, 'International Standards for the Application of the Presumption of Innocence in Criminal Proceedings' (2024) 29(1) Scientific Journal of the National Academy of Internal Affairs 55. doi:10.56215/naia-herald/1.2024.55

18. Khwaileh KM, 'Public Perceptions of Jordan's Cybercrime Law: Protecting Social Media Data and Aligning with International Standards' (2025) 23(1) Pakistan Journal of Life and Social Sciences 2947. doi:10.57239/PJLSS-2025-23.1.00233

19. Ligon NJ, 'Open Trials in the Social Media Age' (2023) 30(3) Virginia Social Policy Review & the Law 288

20. Merle R and Vitu A, *Traité de droit criminel: Procédure pénale* (5th edn, Dalloz 2000) vol 2

21. Obaid AA, The Impact of the Independence of the Judiciary from Conflict in the State of Law: A Comparative Constitutional Study between Arab and International Regulations (Arab Centre for Publishing and Distribution 2018) [in Arabic]

22. Rabinowitz M, '"What Will Become of the Innocent?": Pretrial Detention, the Presumption of Innocence, and Punishment Before Trial' (2023) 7(1) UCLA Criminal Justice Law Review 22. doi:10.5070/CJ87162080

23. Sanders P, 'Protecting the Presumption of Innocence: TikTok Detectives, Misinformation and Private Content Moderation' (2025) 2 Utah Law Review 507

24. Stoyanova M, 'The Impact of Media Publicity on the Presumption of Innocence' in Gies L (ed), *Trial by Media: Participatory Justice in a Networked World* (Palgrave Macmillan 2025) 269. doi:10.1007/978-3-031-80593-6_11

AUTHORS INFORMATION

Tayil Shiyab
PhD (Law), Prof., College of Law, Al Ain University, Al Ain, United Arab Emirates
tayil.sheyab@aau.ac.ae
https://orcid.org/0000-0002-0925-9460
Co-author, responsible for research methodology, data collection, supervising and writing – original draft.

Hakem Alserhan
PhD (Law), Assist. Prof., College of Criminal Justice and Criminal Sciences, Naif Arab University for Security Sciences, Riyadh, Saudi Arabia
halserhan@nauss.edu.sa
https://orcid.org/0000-0001-6834-1248
Co-author, responsible for research methodology, data collection, supervising and writing – original draft.

Mohammad Alkrisheh
PhD (Law), Prof., College of Law, Al Ain University, Al Ain, United Arab Emirates
mohammad.alkrisheh@aau.ac.ae
https://orcid.org/0000-0002-3649-9494
Corresponding author, responsible for research methodology, data collection, writing – review & editing and supervising.

Competing interests: No competing interests were disclosed. Any potential conflict of interest must be disclosed by authors.

Disclaimer: The authors declare that their opinion and views expressed in this manuscript are free of any impact of any organizations.

RIGHTS AND PERMISSIONS

EDITORS

Managing Editor – Mag. Yuliia Hartman. **English Editor** – Julie Bold.
Ukrainian language Editor – Mag. Liliia Hartman.
Arabic language Editor – Mag. Alaa Abdel.

ABOUT THIS ARTICLE

Cite this article
Shiyab T, Alserhan H and Alkrisheh M, 'Legislative Protection for Proper Criminal Justice Procedures Against Publishing on Social Media: A Comparative Analytical Study' (2025) 8(Spec) Access to Justice in Eastern Europe 360-80 <https://doi.org/10.33327/AJEE-18-8.S-c000158>

DOI: https://doi.org/10.33327/AJEE-18-8.S-c000158

Keywords: *criminal justice procedures, social media and criminal trials, fair trial, presumption of innocence, Jordanian penal code, cybercrime law, comparative criminal law, judicial independence.*

DETAILS FOR PUBLICATION

Date of submission: 05 Oct 2025
Date of acceptance: 03 Nov 2025
Online First Publication: 10 Dec 2025
Last Published: 30 Dec 2025
Whether the manuscript was fast tracked? - No
Number of reviewer report submitted in first round: 2 reports (2 external reviewers)
Number of revision rounds: 1 round with conditional acceptance

Technical tools were used in the editorial process:
Plagiarism checks - Turnitin from iThenticate https://www.turnitin.com/products/ithenticate/
Scholastica for Peer Review https://scholasticahq.com/law-reviews

AI DISCLOSURE STATEMENT

The corresponding author confirmed that the manuscript was written entirely by the authors. AI tools were used exclusively for spelling, grammar, and stylistic refinement. No generative AI tools were used to create original content, research ideas, interpretations, analyses, or conclusions.

АНОТАЦІЯ УКРАЇНСЬКОЮ МОВОЮ

Тематичне дослідження

ЗАКОНОДАВЧИЙ ЗАХИСТ НАЛЕЖНИХ ПРАВОВИХ ПРОЦЕДУР У КРИМІНАЛЬНОМУ ПРОЦЕСІ ВІД ПУБЛІКАЦІЙ У СОЦІАЛЬНИХ МЕРЕЖАХ: ПОРІВНЯЛЬНО-АНАЛІТИЧНЕ ДОСЛІДЖЕННЯ

Таїл Шіяб, Хакем Альсерхан та Мохаммад Алькрішех*

АНОТАЦІЯ

Вступ. *В епоху цифрових технологій платформи соціальних мереж стали потужним простором для залучення громадськості до питань кримінального судочинства, часто впливаючи на уявлення про вину та невинуватість. У цьому контексті правова система Йорданії стикається зі все більшими викликами щодо захисту цілісності судових процесів від наслідків передчасних або упереджених онлайн-публікацій. Це дослідження вивчає, наскільки йорданське законодавство захищає ключові процесуальні гарантії, а саме презумпцію невинуватості, конфіденційність розслідувань та неупередженість судової влади, від впливу контенту соціальних мереж та публічних коментарів щодо кримінальних справ.*

Методи. *У дослідженні використовується доктринальна та порівняльно-правова методологія. У ньому аналізуються відповідні положення Кримінального кодексу Йорданії, Закон про кіберзлочинність № 17 від 2023 року та Закон про телекомунікації № 13 від 1995 року, а також відповідні процесуальні закони. Дослідження також спирається на порівняльний аналіз законодавства Франції та Еміратів, щоб оцінити ступінь відповідності йорданського законодавства міжнародним стандартам захисту кримінальних проваджень від втручання ЗМІ та упередженості в Інтернеті.*

Результати та висновки. *Результати дослідження показують, що законодавство Йорданії ще не забезпечує достатньо чітких або вичерпних гарантій для запобігання впливу на правосуддя, що здійснюється через тиск громадськості чи ЗМІ. Дослідження висвітлює значні прогалини щодо захисту слідчої таємниці та неупередженості судових органів. Відповідно, було подано рекомендації запровадити чіткі кримінальні положення, щоб заборонити передчасну публікацію матеріалів розслідувань, захистити прокурорів та суддів від стороннього впливу та посилити покарання за несанкціоноване розкриття інформації. Ці реформи зміцнять правосуддя та незалежність системи кримінального судочинства, зберігаючи при цьому баланс у дотриманні свободи слова в цифровій сфері.*

Ключові слова: *процедури кримінального судочинства, соціальні мережі та кримінальні судові процеси, справедливий суд, презумпція невинуватості, Кримінальний кодекс Йорданії, Закон про кіберзлочинність, порівняльне кримінальне право, незалежність судової влади.*

Access to Justice in Eastern Europe
ISSN 2663-0575 (Print) ISSN 2663-0583 (Online)
Journal homepage http://ajee-journal.com

ABSTRACT IN ARABIC

حالة الدراسة

الحماية التشريعية لإجراءات العدالة الجنائية من النشر على وسائل التواصل الاجتماعي: دراسة تحليلية مقارنة

تيل شياب وحاكم السرحن ومحمد الكريشية*

الملخص

الخلفية: *في العصر الرقمي، أصبحت منصّات التواصل الاجتماعي فضاءات مؤثرة لإشراك الجمهور في قضايا العدالة الجنائية، وغالبًا ما تسهم في تشكيل تصورات الرأي العام بشأن الإدانة أو البراءة. وفي هذا السياق، يواجه النظام القانوني الأردني تحديات متزايدة في حماية نزاهة الإجراءات القضائية من الآثار السلبية للنشر الإلكتروني المبكر أو المنحاز. وتتولى هذه الدراسة بحث مدى قدرة التشريعات الأردنية على صون الضمانات الإجرائية الرئيسة، ولا سيما قرينة البراءة، وسرية التحقيق، وحياد السلطة القضائية، من تأثير المحتوى المنشور على وسائل التواصل الاجتماعي والتعليقات العامة المرتبطة بالقضايا الجنائية.*

المنهجية: *تعتمد الدراسة منهجًا قانونيًا تحليليًا مقارنًا، حيث تقوم بتحليل نقدي لأحكام قانون العقوبات الأردني، وقانون الجرائم الإلكترونية رقم 17 لسنة 2023، وقانون الاتصالات رقم 13 لسنة 1995، إلى جانب القوانين الإجرائية ذات الصلة. كما تستخلص الدراسة رؤى مقارنة من التشريعين الفرنسي والإماراتي بهدف تقييم مدى اتساق القانون الأردني مع المعايير الدولية المخصصة لحماية الإجراءات الجنائية من التدخل الإعلامي والتحيز الرقمي.*

نتائج واستنتاجات: *وتكشف النتائج عن أن التشريعات الأردنية لا تزال تفتقر إلى الضمانات الصريحة والشاملة الكفيلة بمنع تأثير الضغط الإعلامي أو المجتمعي على سير العدالة. وتبرز الدراسة وجود فجوات واضحة تتعلق بحماية سرية التحقيقات وضمان حياد السلطات القضائية. وبناءً على ذلك، توصي الدراسة بسنّ نصوص جنائية صريحة تحظر نشر المواد التحقيقية قبل أوانها، وتكفل حماية أعضاء النيابة العامة والقضاة من التأثيرات غير المشروعة، وتعمل على تشديد العقوبات المترتبة على أي إفصاح غير مصرح به. من شأن هذه الإصلاحات تعزيز عدالة واستقلالية النظام الجنائي، مع الحفاظ على التوازن المطلوب بين هذه الحماية وحرية التعبير في الفضاء الرقمي.*

For further information, please contact
Access to Justice in Eastern Europe, Publishing House 'Academic Insights Press',
Emil-Bertalanffy-Weg 1, Unterach am Attersee, 4866, Austria
https://academicinsightspress.com

Postal information
Publishing House 'Academic Insights Press', Emil-Bertalanffy-Weg 1, Unterach am Attersee, 4866, Austria.

Permissions
For information on how to request permissions to reproduce articles from this journal, please visit our web-site and contact Editor-in-Chief.

Advertising
Inquiries about advertising and inserts should be addressed to info@ajee-journal.com.

Disclaimer
Statements of fact and opinion in the articles and notes in *Access to Justice in Eastern Europe* are those of the respective authors and contributors and not of *Access to Justice in Eastern Europe*.
No *Access to Justice in Eastern Europe* issue makes any representation, expressed or implied, regarding the accuracy of the material in this journal and cannot accept any legal responsibility or liability for any errors or omissions that may be made. The reader should make his/her evaluation as to the appropriateness or otherwise of any experimental technique described.

Media identifier R30-00995
The print run of this issue is 200 copies

Офіційна назва зареєстрованого друкованого медіа - *Доступ до правосуддя в Східній Європі, Access to Justice in Eastern Europe*
Порядковий номер випуску друкованого медіа та дату його виходу у світ - № 4(Special) за 2025 рік
Тираж випуску - 200 екз.
Ідентифікатор друкованого медіа у Реєстрі - R30-00995

Cover-list by **Alona Hrytsyk**

Publishing House 'Academic Insights Press'
Emil-Bertalanffy-Weg 1, Unterach am Attersee, 4866, Austria
https://academicinsightspress.com

Access to Justice in Eastern Europe (AJEE) is a peer-reviewed, Open Access journal. The main aim of AJEE is to provide a forum for the discussion of topical issues related to judiciary and civil procedure reforms, as well as the sharing of research results on access to justice developments in Eastern European countries. This scope has been chosen due to the unique combination of post-socialist legal doctrine's legacy, Western influence, and the legislative approximation to EU law.

We also welcome and encourage submissions from other regions that explore comparative perspectives or offer insights into global trends and developments related to access to justice.

EELRC
East European Law Research Center

Akademische Einblicke Verlag