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

Research Article

TRANSPARENCY IN THE LABYRINTHS OF THE EU AI ACT: SMART OR DISBALANCED?

Gintare Makauskaite-Samuole

ABSTRACT

Background: *Complete transparency in artificial intelligence is impossible to achieve.*[1] *In the interdependent technological context, the scope of artificial intelligence transparency and the logic behind the values that outweigh transparency are unclear. Legislation on artificial intelligence, such as the European Union Artificial Intelligence Act (hereinafter the EU AI Act), tries to define the true meaning and role of AI transparency.*

Methods: *The author applies doctrinal research and comparative analysis methods to assess AI transparency in the EU AI Act; a framework of distinct transparency zones is established. Doctrinal research helps to define the scope of transparency obligations and examine their limitations and interaction within the EU AI Act, while comparative analysis highlights inconsistencies, such as an unexplained difference between transparency duties in distinct zones or different requirements for open source and proprietary AI.*

Results and conclusions: *The findings reveal a fragmented and uneven framework of artificial intelligence transparency in the EU AI Act, shaped by many exemptions, exceptions, derogations, restrictions, and other limitations. The zero-transparency zone (established by Article 2) is too broad, with much discretion given to stakeholders. In contrast, the basic transparency zone (set by Article 50) is too narrow, posing risks to fundamental human rights. The next zone, the moderate transparency zone (Chapter V), struggles with responsibility sharing between AI providers and downstream deployers. Meanwhile, the high transparency zone (provided in Chapter III) privileges law enforcement. Lastly, the hybrid transparency zone highlights complications in managing interactions between different risk-level AI systems.*

1 Mona Sloane and others, 'Introducing Contextual Transparency for Automated Decision Systems' (2023) 5 Nature Machine Intelligence 188, doi:10.1038/s42256-023-00623-7.

The author concludes that the EU AI Act is progressive but needs more fine-tuning to function as a coherent and solid transparency framework. The scales between public interest in artificial intelligence transparency, individual and societal rights, and legitimate interests risk being calibrated post-factum.

1 INTRODUCTION

The European Union (EU) faces distinct challenges in the field of artificial intelligence (AI). Member States have been slow to reach the targets set by the Digital Agenda,[2] and only a few European AI startups, such as Mistral AI, can compete with their non-European counterparts, making Europe a *terra nullius* market for non-European AI companies. According to estimates from the European Commission, 70% of AI systems in the EU pose minimal risk[3] and are not bound by transparency obligations under the EU AI Act.[4] This limited outlook raises the question of whether humans will control AI or AI will control humans.

Transparency, as outlined in Recital 9 of the EU AI Act, is one of its rationales "to strengthen the effectiveness of such existing rights <..> by establishing <..> obligations <..> in respect of transparency".[5] Its primary objective is to establish a balance between innovation and the protection of human rights. However, this regulatory compromise inherently reflects the EU's fundamental values. The problem is that the regulatory compromise is often a dangerous trade-off.

The challenge lies in the extent to which transparency is limited by legislative choices—determining what is visible and where scrutiny is directed. While the lack of AI transparency may stimulate innovation development in the short term, in the long run, inadequate transparency may deepen the gap between expectations and reality. This imbalance could have a wide effect on society and human rights.

2 European Commission, 'Second Report on the State of the Digital Decade calls for Strengthened Collective Action to Propel the EU's Digital Transformation: Press Release' (2 July 2024) <https://ec.europa.eu/commission/presscorner/detail/en/ip_24_3602> accessed 5 August 2024.

3 European Commission, *Commission Staff Working Document: Impact Assessment Accompanying the Proposal for a Regulation of the European Parliament and of the Council Laying down Harmonized Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*, SWD/2021/84 final (21 April 2021) <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52021SC0084> accessed 5 August 2024.

4 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) PE/24/2024/REV/1 [2024] OJ L 235/1 <http://data.europa.eu/eli/reg/2024/1689/oj> accessed 5 August 2024.

5 Luca Holst and others, 'The Impact of the EU AI Act's Transparency Requirements on AI Innovation' (19th International Conference on Wirtschaftsinformatik (WI), Würzburg, Germany, September 2024) <https://eref.uni-bayreuth.de/id/eprint/90313/> accessed 17 November 2024.

EU AI Act's adoption for Member States is costly,[6] even if the effect is partially complemented by the compliance of GDPR and other digital acquis. Uneven and slow digitalisation within the EU calls for simplification, verticalisation, or deregulation of AI.[7] Faced with variable and stringent "Brussels effect" [8], Eastern European candidate countries double-track laws[9] for future compliance, end in weak enforcement or misuse exemptions for nondemocratic reasons.[10] A solid transparency framework would minimise the mentioned side effects and bring internal and external clarity about the EU's values.

Systemic, generalised analysis of AI transparency in light of substantive limitations and constraints is not often discussed in the academic literature. Scholars concentrate on specific aspects of transparency, such as trust and accountability, high-risk AI systems, explainability, or suitability of a risk-based approach to protecting rights. However, AI transparency requires a more comprehensive, high-detail approach from the perspective of the limitations of transparency frameworks. The article argues that the role of AI transparency is shadowed and overcomplicated by the tiered risk management framework in the EU AI Act; the scope of AI transparency is unevenly distributed. AI transparency, inter alia, depends on limitations and constraints. By analysing them, the author assesses if artificial transparency in the EU AI Act is sufficient to address the current needs of a digital European future.

2 METHODOLOGY

The research focuses on the provisions for AI transparency in the EU AI Act. The methodology includes a doctrinal research method and comparative analysis of transparency regulation in the EU AI Act.

The first part of the paper provides an overview of AI transparency particularities in the EU AI Act. Two general aspects are considered briefly using doctrinal research and comparative analysis. First, AI transparency's role and true meaning are evaluated in contrast to the already established governmental transparency framework. It became evident that the tailored form of transparency operates within a complex tiered risk management

6 Meeri Haataja and Joanna J Bryson, 'What Costs Should We Expect from the EU's AI Act?' (*Center for Open Science*, 27 August 2021) SocArXiv 8nzb4, doi:10.31235/osf.io/8nzb4.

7 Mario Draghi, *The Future of European Competitiveness: Part A: A Competitiveness Strategy for Europe* (European Commission 2024).

8 Charlotte Siegmann and Markus Anderljung, *The Brussels Effect and Artificial Intelligence: How EU Regulation Will Impact the Global AI Market* (Center for the Governance of AI 2022) 3-4.

9 Ministry of Digital Transformation of Ukraine, *The White Paper on Artificial Intelligence Regulation in Ukraine: Vision of the Ministry of Digital Transformation of Ukraine : Version for Consultation* (Ministry of Digital Transformation of Ukraine 2024) 17.

10 Amnesty International, *"A Digital Prison": Surveillance and the Suppression of Civil Society in Serbia* (Amnesty International Ltd 2024).

framework, where stakeholder roles and responsibilities are blurred. Additionally, the application of transparency rules is not fine-tuned. Following this, the paper reviews the limitations of AI transparency, highlighting numerous constraints that stem from various provisions yet lack consolidation.

The second part of the paper reviews the scope constraints of AI transparency. The author proposes a novel framework of transparency zones, which helps differentiate the AI transparency framework from the risk management framework. This approach allows for an individual and collective assessment of each zone, offering a clearer understanding of how transparency provisions function and interact. Transparency zones are reviewed to cover data, algorithm and output transparency, with substantial elements compared across zones. The paper further evaluates the gradual increase in transparency obligations zone by zone, identifying the target audience and weighing the arguments pro and contra transparency values. The study assesses whether transparency obligations are balanced and whether they impose an excessive burden on one side. Finally, transparency obligations are assessed to check if they are feasible and non-contradictory within a developing AI landscape. The research is concluded with final remarks.

3 OVERVIEW OF AI TRANSPARENCY IN THE EU AI ACT

3.1. Particularities of AI transparency in the EU AI Act11

AI transparency is a broad "umbrella" concept that is contextualised for different audiences and environments.[12] The broad concept of AI transparency, as used in transparency frameworks,[13] overlaps with other AI system properties, like explainability and interpretability, to address ethical and societal concerns.[14] This article uses the broad concept of AI transparency to correspond to different environments where its limitations occur.

11 Artificial Intelligence Act (n 3) preamble, para 64.

12 Anastasiya Kiseleva, Dimitris Kotzinos and Paul De Hert, 'Transparency of AI in healthcare as a multilayered system of accountabilities: between legal requirements and technical limitations' (2022) 30(5) Frontiers in artificial intelligence 7-8, doi:10.3389/frai.2022.879603.

13 Kashyap Haresamudram, Stefan Larsson and Fredrik Heintz, 'Three levels of AI transparency' (2023) 56(2) Computer 93, doi:10.1109/MC.2022.3213181; Md Tanzib Hosain and others, 'Path to Gain Functional Transparency in Artificial Intelligence with Meaningful Explainability ' (2023) 3(2) Journal of Metaverse 166, doi:10.57019/jmv.1306685; Luca Nannini, 'Habemus a Right to an Explanation: so What? – A Framework on Transparency-Explainability Functionality and Tensions in the EU AI Act' (2024) 7(1) Proceedings of the AAAI / ACM conference on AI, Ethics, and Society 1023, doi:10.1609/aies.v7i1.31700.

14 ISO/IEC FDIS 12792 Information Technology - Artificial Intelligence - Transparency Taxonomy of AI Systems (ISO/IEC DIS 12792:2024) (ISO/IEC 2024) 7.

AI transparency generally aims to make regulated AI systems' processes, decisions and processes visible and understandable throughout their lifecycle. Visibility includes both short-term and long-term processes and results. Understandability upgrades AI transparency from mere technical visibility by ensuring a specific level of comprehension for the audience, enabling informed choice, as provided in the OECD Principles on AI.[15]

The EU AI Act establishes a tiered risk management framework on which, based on the reading of Recital 27, the transparency framework largely depends. However, the risk management framework's dominance and shortcomings—criticised for being too arbitrary, not based on objective data,[16] and lacking dynamic risk-benefit analysis—[17] suppresses the transparency framework's visibility. Unsurprisingly, existing classifications, such as technical and protective transparency or rights-enabling, review-enabling, and decision-enabling transparency,[18] show that transparency is sometimes reduced to an operational attribute of the risk management framework.

In the EU AI Act, AI transparency does not benefit from the principle of maximum disclosure or transparency by default, as governmental transparency does. Instead, it is a targeted and tailored form of transparency. Targeted transparency has several implications for AI regulation. First, targeted transparency means limited scope: not all AI systems are subject to obligatory transparency requirements due to the narrow definition of AI, AI system, risk, and other relevant terms. Second, targeted transparency affects the balancing of interests. Third, it also implies the application of the proportionality principle, meaning that the scope of transparency regulation and the subsequent application of transparency requirements is subject to the necessity and proportionality of regulatory intervention.

In addition, AI transparency under the EU AI Act is restrictive, somewhat inadequate, and operates through separate communication "lanes". The Act establishes a risk-based tiered compliance framework involving AI providers, deployers, and public and supervisory institutions. Public transparency is limited, with the public playing a passive role and lacking participatory rights typically granted in other regulatory contexts.[19] Responsibility is concentrated in the hands of supervisory institutions and AI providers (and, in some cases, deployers) via organisational and expert transparency. While AI developers and deployers

15 OECD, *Recommendation of the Council on Artificial Intelligence*, OECD/LEGAL/0449 (OECD Legal Instruments 2025).

16 Martin Ebers, 'Truly Risk-Based Regulation of Artificial Intelligence: How to Implement the EU's AI Act' [2024] European Journal of Risk Regulation 7-10, doi:10.1017/err.2024.78.

17 Henry Fraser and José-Miguel Bello y Villarino, 'Acceptable Risks in Europe's Proposed AI Act: Reasonableness and Other Principles for Deciding How Much Risk Management Is Enough' (2024) 15(2) European Journal of Risk Regulation 445, doi:10.1017/err.2023.57.

18 Holst and others (n 5).

19 Kostina Prifti and others, 'From Bilateral to Ecosystemic Transparency: Aligning GDPR's Transparency Obligations with the European Digital Ecosystem of Trust' in Simone Kuhlmann and others (eds), *Transparency or Opacity: A Legal Analysis of the Organization of Information in the Digital World* (Nomos 2023) 135-9, doi:10.5771/9783748936060-115.

have some transparency obligations towards the individuals, most transparency measures—such as documentation and auditing—are directed at supervisory institutions. Furthermore, knowledge sharing between AI providers and downstream deployers is designed with a high degree of discretion for AI providers.

Some framework inconsistency is preprogrammed in the timing when transparency measures are applicable. Unlike others, the EU AI Act concentrates on transparency measures on the launch of the AI system or entrance to the market, with the risk of transparency being assessed *post-factum* and converted into a larger scale risk. Focusing on a specific point may lead to blanket compliance. The ideation, testing, and even post-mortem stages require transparency—especially considering the impact of initial errors or reuse of system components after discontinuation.

EU lawmakers have taken the pragmatic approach to avoid overburdening both themselves and AI market participants. Timing and scope constraints illustrate this point. The application of transparency rules is narrow. The involvement of multiple loosely connected stakeholders further diffuses accountability for AI non-transparency. Meanwhile, the public—arguably the least informed participant in the AI ecosystem—remains passive and lacks participatory rights. These factors, taken together, signal a highly granular approach to AI transparency. Adding to this complexity, the EU AI Act establishes numerous transparency limitations, further shrinking the scope of transparency obligations.

Limitations of AI Transparency

The EU AI Act provides various limitations on transparency: exemptions, exceptions, derogations and restrictions.

First, some *exemptions* in Article 2 of the EU AI Act prevent its application. The logical line between them is not consistent. They include exemptions in *scope*, such as national security, scientific research, and personal use; in *time*, such as pre-market development; and in *type*, like free open-source systems.

Second, some *exceptions* in Article 50 release AI providers and deployers from specific transparency obligations. Exceptions are cumulative (like law enforcement, which releases from several transparency duties, raising questions about why law enforcement is so privileged) and single, like AI assistive role.

Thirdly, some *restrictions* arise from normative competition related to the interaction with other legislation. On one side, the general rule in Article 50 states that transparency duties for certain AI systems are "without prejudice" to other EU or national law transparency obligations; regulatory complementarity and harmonising nature are elaborated in Recital 9. On the other, there are hints of prioritised obligations within the text. They include the confidentiality duty of supervisory institutions set in Article 78 of the EU AI Act, which coincides with typical exceptions to access official information (such as IP rights, trade secrets, audits, national security, etc.). They also include rules for respect of privacy and

personal data protection (Recital 69 of the EU AI Act), the right to clear and meaningful explanations for decisions having legal effects made by high-risk AI systems (Article 86); bias removal vs. sensitive personal data (Article 10); respect for specific substantial public interests in AI regulatory sandboxes (Article 59), as well as the duty to conduct fundamental right impact assessments in high-risk systems. Without the general principle of pro-disclosure, norm competition is too complicated to be free from tensions over what takes priority – public interest in AI transparency or confidentiality.

Fourth, the EU AI Act applies proportionality to AI transparency rules. Proportionality is predetermined, without dynamic balance. The first example is the tiered risk framework, with different transparency obligations corresponding to risk level. Higher-risk AI systems have more extensive transparency obligations than low-risk. AI regulatory sandboxes have lighter transparency obligations compared to standard AI systems. Similarly, in *derogations*, microenterprises and smaller organisations are subject to simplified transparency obligations (Article 17, 63); inconsistently, the risk is not the dominant determinant in this case. In addition, synthetic content generated by AI as part of evidently artistic, creative, satirical, fictional, or analogous work or program is subject to lighter requirements (Article 50), predetermining a balance with freedom of arts.

To sum up, the EU AI Act's approach to AI transparency limitations is complex and fragmented. The transparency framework lacks a systemic pro-disclosure foundation needed for robustness and future-proofing, making it difficult to measure and compare. The limitations are crafted as single or cumulative—sometimes separate, sometimes layered—justified by the strict necessity for regulatory intervention. While the protected values counterbalanced against pro-disclosure are context-specific, there is a lack of a logical explanation for why certain values were selected over others or why not all fundamental rights and freedoms were considered. In this complex labyrinth of limitations, the meaning of transparency risks being lost. However, an extensive scope of transparency obligations, precisely calibrated to specific types or uses of AI systems, could mitigate some of these negative effects to a certain extent. The following section examines transparency zones to assess this possibility.

4 TRANSPARENCY ZONES

Importantly, the transparency framework within the EU AI Act is not limited to the risk management framework. Instead, it acts as a spectrum shaped by the Act's risk-based approach, soft law principles, proportionality considerations, and specific limitations.

For practical purposes—such as assessing transparency obligations and enabling comparability with other frameworks—several transparency zones can be distinguished within the EU AI Act. These zones differ in the scope of transparency obligations, limitations and procedural safeguards. They include zero transparency, basic transparency, moderate transparency, and high-risk zones. The scope of

obligations and derogations reflect the width of transparency, while its depth is determined by exceptions and restrictions.

The participants of the transparency framework include AI providers (platform providers or product or service providers), relevant authorities (policymakers and regulators at the EU and national level), AI partners (AI system integrators, data providers, AI evaluators, and AI auditors), AI subjects (data subjects and other subjects), AI customers (AI users) and AI producers (AI developers).[20] Indirect participants, such as hosting providers, are not included.[21] Sometimes, the stakeholder groups become intertwined due to the role of a stakeholder. For example, a deployer may transition into a provider due to extensive fine-tuning. As a result, transparency mechanisms extend beyond primary stakeholders like AI providers to encompass second-line stakeholders, such as developers of AI integrators.

The target audience is narrower and includes four groups: users, researchers and developers, integrators and providers, and regulators and third-party auditors.[22] Transparency enables them to fulfil their roles. For example, integrators can investigate incidents, regulators can assess performance, researchers can test and validate AI models, and users can make informed decisions.

Transparency measures can be grouped into three main types: data transparency, which involves disclosure of data sources, data processing, data security, etc.; algorithmic transparency, which entails revealing algorithms or model architectures; and outcome transparency, which focuses on disclosing AI system results and decisions.[23] However, the burden on AI stakeholders is uneven. A tiered system of transparency measures is established, with the most important attributes of AI transparency being interpretability, explainability, and traceability.[24]

The EU AI Act defines four transparency zones with varying levels. The first zone, zero transparency, is based on voluntary disclosure. The second zone, basic transparency, centres on AI awareness duty. The third zone, moderate transparency, requires several measures regarding risks, copyright, and data. The fourth and high-risk transparency zone introduces extensive transparency measures for proper accountability, safety, traceability, and control.

20 'ITI Policy Principles for Enabling Transparency of AI Systems' (*Information Technology Industry Council*, September 2022) <https://www.itic.org/policy/artificial-intelligence/iti-policy-principles-for-enabling-transparency-of-ai-systems> accessed 18 June 2024.

21 Thalia Khan and Madhulika Srikumar, 'Developing General Purpose AI Guidelines: What the EU Can Learn from PAI's Model Deployment Guidance' (*Partnership on AI*, 26 November 2024) <https://partnershiponai.org/developing-general-purpose-ai-guidelines-what-the-eu-can-learn-from-pais-model-deployment-guidance/> accessed 18 December 2024.

22 ITI Policy Principles (n 20).

23 Yinuo Geng, 'Transparency for What Purpose?: Designing Outcomes-Focused Transparency Tactics for Digital Platforms' (2024) 16(1) Policy & Internet 83, doi:10.1002/poi3.362.

24 Jessica Kelly and others, 'Navigating the EU AI Act: A Methodological Approach to Compliance for Safety-Critical Products' (*arXivLabs*, 26 March 2024) arXiv:2403.16808 [cs.AI], doi:10.1109/CAI59869.2024.0017

4.1. Zero Transparency Zone

The zero-transparency zone (Articles 2.3, 2.4, 2.6, 2.7, 2.8, 2.9, 2.10, 2.12) includes areas covered by sectoral exemptions and minimal-risk AI systems, allowing major stakeholder groups—individuals, experts, businesses, and governments—to experiment in dedicated fields. Individuals are entitled to a free-hand approach in personal, small-scale projects, experts can conduct research, businesses can innovate until they introduce AI systems to the market or offer open-source systems, and governments can act in national security, military, and defence. In this context, transparency is outweighed by the overriding public interest (exemptions) or private interest (minimal-risk AI systems).

At the same time, zero transparency reflects the EU's policy to promote AI proliferation in areas like national security, military and defence, research, personal or prior market experimentation, free and open-source systems, and minimal risk systems. This zone becomes a playground for innovation and a competitive advantage for the EU market. However, leaving risks in the zero-transparency zone unaddressed—especially in high-risk domains like autonomous weapons—fails to ensure full regulatory comprehensiveness. While some risks could benefit from control or, at least, some visibility, the zero-transparency zone effectively limits the scope of oversight.

The balance test between competing rights and interests depends on the context of the intended use of the AI system. However, it lacks a pro-transparency focus as it does not incorporate harm or public interest override tests. The only safeguard is the sole purpose test, which applies in cases like national security, though not universally.

Zero Transparency Zone	Pro Transparency Values	Contra Transparency Values
National security, personal activities, international cooperation, open-source systems, research, and testing, etc.	-	Protection of personal data Respect for private life Freedom of expression Right to liberty and security Right to a fair trial Freedom to conduct business

Two significant factors shape the scope of the zero-transparency zone: the broad interpretation of exemptions and the sole purpose test.

National security (Article 2.3) is the sole responsibility of individual EU member states and is directly excluded from the EU AI Act's application. This contrasts with the Framework Convention on Artificial Intelligence, which treats national security as an optional exemption while still requiring adherence to human rights standards and encouraging the

voluntary application of high-risk AI system standards.[25] It is the broad discretion of the states to determine national security's conceptual limits flexibly; the normative ambiguity is preprogrammed. Due to the free-hand approach, AI could become the weapon of first resort in future conflicts. [26]

The sole purpose test requires a direct link to national security. Even temporary use of national security AI systems for other purposes—including humanitarian purposes and public security—falls under the scope of the EU AI Act. In cases of dual-use, the AI Act applies to non-national security uses, but the concepts still need to be delineated based on the member states' discretion. Based on the CJEU case law on data retention and national security concerns (examples here[27] and here[28]), proportionality and necessity must be assessed, even though national security remains the sole responsibility of individual states.

These challenges are further compounded by the increasing reliance on private bodies to develop intelligent technologies for national security.[29] Industry trends indicate that initiatives with adaptable solutions often come from the private sector, not the state. Additionally, national security bodies frequently rely on standard third-party AI tools and functionalities,[30] and many non-adversarial AI systems are inherently dual-use by nature due to the universality of the algorithm.[31] This dual-use reality creates a dilemma: applying transparency and compliance requirements to non-military applications while potentially disadvantaging EU actors against adversaries. Consequently, tailored exclusive-use national security systems may proliferate as a result.

Other exemptions in the zero-transparency zone, like scientific research and development (Article 2.8) and minimal-risk AI systems (Articles 2.10, 2.12), also benefit from a broad interpretation of concepts and the sole purpose test. The potential risk of unfair use of the

25 Rosamund Powell, 'The EU AI Act: National Security Implications' (*CETaS Explainers*, 31 July 2024) <https://cetas.turing.ac.uk/publications/eu-ai-act-national-security-implications> accessed 18 December 2024.

26 Emre Kazim and others, 'Proposed EU AI Act–Presidency Compromise Text: Select Overview and Comment on the Changes to the Proposed Regulation' (2023) 3 AI Ethics 382, doi:10.1007/s43681-022-00179-z.

27 *Digital Rights Ireland Ltd v Minister for Communications, Marine and Natural Resources and Others and Kärntner Landesregierung and Others* Joined Cases C 293/12 and C 594/12 (CURIA (Grand Chamber), 8 April 2014) <https://curia.europa.eu/juris/liste.jsf?num=C-293/12> accessed 18 December 2024.

28 *Tele2 Sverige AB v Post- och telestyrelsen and Secretary of State for the Home Department v Tom Watson and Others* Joined Cases C-203/15 and C-698/15 (CURIA (Grand Chamber), 21 December 2016) <https://curia.europa.eu/juris/liste.jsf?num=C-203/15> accessed 18 December 2024.

29 Powell (n 25).

30 ibid.

31 Rosanna Fanni, 'Why the EU Must Now Tackle the Risks Posed by Military AI' (*CEPS*, 8 June 2023) <https://www.ceps.eu/why-the-eu-must-now-tackle-the-risks-posed-by-military-ai/> accessed 18 December 2024.

scientific research and development exemption for the end intent of commercialisation is observed, [32] or when publicly funded research ends with no commercialisation but is used. [33]

The personal use exemption is also contested in the context of AI liability. Non-professionals without expertise are subject to a lower standard of conduct than professionals and are not required to abide by the same standard of care as professionals.[34] As the presumption of causation is not applicable for non-professionals in the draft AI Liability Directive,[35] and there is no transparency data to rely on, it is complicated for national judges to solve such cases.

Furthermore, the exemption for minimal-risk AI systems further expands the zero transparency zone, leaving most AI systems (70%)[36] unregulated, with some areas even preempting national national law.[37] Proponents of controllable AI advocate for applying transparency to all AI systems;[38] include XAI[39] and proportional transparency measures.[40]

What does this mean for the enforcement of the EU AI Act? Vague definitions and broad exemptions for national security, minimal risk, and research and development leave large areas either unregulated or governed by different standards. Major stakeholder groups face a dual-standard problem, which may discourage or impede the enforcement of stricter transparency standards unless alternative enforcement mechanisms are introduced.

The public is the least privileged regarding the right to know, facing risks of misuse, such as social engineering. The under-regulation of potential harms requires proactive monitoring and future changes to avoid systemic risks for human rights protection. Given these challenges, the scope of the basic transparency zone—which establishes minimal disclosure requirements—is of critical importance.

32 Kazim and others (n 26) 386.

33 Liane Colonna, 'The AI Act's Research Exemption: A Mechanism for Regulatory Arbitrage?' in Eduardo Gill-Pedro and Andreas Moberg (eds), *YSEC Yearbook of Socio-Economic Constitutions 2023: Law and the Governance of Artificial Intelligence* (Springer Cham 2023) 59, doi:10.1007/16495_2023_59.

34 Cristina Frattone, 'Reasonable AI and Other Creatures: What Role for AI Standards in Liability Litigation?' (2022) 1(3) Journal of Law, Market & Innovation 38, doi:10.13135/2785-7867/7166.

35 ibid 15-55.

36 Marc P Hauer and others, 'Quantitative Study About the Estimated Impact of the AI Act' (*arXivLabs*, 29 March 2023) arXiv:2304.06503 [cs.CY], doi:10.48550/arXiv.2304.06503.

37 Ida Varošanec, 'On the Path to the Future: Mapping the Notion of Transparency in the EU Regulatory Framework for AI' (2022) 36(2) International Review of Law, Computers & Technology 104, doi:10.1080/13600869.2022.2060471.

38 Peter Kieseberg and others, 'Controllable AI - An Alternative to Trustworthiness in Complex AI Systems?' in Andreas Holzinger and others, (eds), *Machine Learning and Knowledge Extraction: 7th IFIP TC 5, TC 12, WG 8.4, WG 8.9, WG 12.9 International Cross-Domain Conference, CD-MAKE 2023, Benevento, Italy, 29 August – 1 September 2023* (LNCS 14065, Springer Cham 2023) 1, doi:.1007/978-3-031-40837-3_1.

39 Anetta Jedličková, 'Ethical Considerations in Risk Management of Autonomous and Intelligent Systems' (2024) 14(1-2) Ethics & Bioethics 91, doi:10.2478/ebce-2024-0007.

40 Claudio Novelli and others, 'AI Risk Assessment: A Scenario-Based, Proportional Methodology for the AI Act' (2023) 3 Digital Society 13, doi:10.1007/s44206-024-00095-1.

4.2. Basic Transparency Zone

The general obligation of providers and deployers of all AI systems within the scope of the EU AI Act is to ensure AI literacy, as regulated in Article 4. Providers and deployers must ensure that their staff and others involved in operating or using AI systems on their behalf achieve adequate AI literacy. This involves technical knowledge, experience, education, and training, as well as the specific context in which the AI systems will be used and the needs of end-users. The enforcement of this obligation, particularly in terms of transparency, largely depends on regulatory oversight of AI-related documentation, such as training data.

As the primary transparency obligation in a basic transparency zone, AI disclosures draw a line between the human world and the synthetic domain. On one hand, they require disclosure to natural persons when they are exposed to AI or AI output. On the other hand, they require disclosure for machines, indicating that the content is synthetic. However, compliance with transparency obligations does not automatically render an AI system or its output lawful (Recital 137).

Article 50 of the EU AI Act sets out five disclosures, including the disclosure of human interaction with AI, technical watermarking of synthetic content, disclosure of emotion recognition/biometric categorisation systems, disclosure of deepfake content, and disclosure of AI-generated text for the public interest. All these disclosures fall under the scope of AI outcome transparency, not including data or algorithmic transparency layers. The scope of AI disclosures varies. AI HLEG principles suggest it should include "making humans aware that they communicate or interact with an AI system, as well as duly informing deployers of the capabilities and limitations of that AI system and affected persons about their rights" (Recital 27). AI disclosures for low-risk AI systems do not include details about the provider of the AI system, logic beneath the AI algorithms, instructions of use, rights, responsibilities, and remedies, making specific AI documentation available publicly, registering in AI systems registers, etc. The burdensome effect of broader transparency is presumed in that way.

Article 50 further mandates that AI interaction or exposure disclosures to natural persons in specific-risk AI systems. Relevant information must meet the criteria of timeliness, clarity, distinguishability, and accessibility (Article 50.5). Clarity requires that disclosures be high-level and easily understood by the general public, while distinguishability ensures that the public notices. Accessibility is most likely to be interpreted in the narrow sense,[41] but not implying comprehensibility and usability for the intended user. As there are no specific requirements for explanations of system operation, limitations, or user expectations, the discretion to indicate that AI may make mistakes or accuracy metrics rests upon the provider.

41 Directive (EU) 2019/882 of the European Parliament and of the Council of 17 April 2019 on the Accessibility Requirements for Products and Services (Text with EEA Relevance) PE/81/2018/REV/1 [2019] OJ L 151/70.

The values to be weighed in the balance test depend on the context of the AI system's intended use and are not limited to those directly mentioned in the EU AI Act. However, some of them are mentioned in the exceptions to AI disclosure (such as Article 50.3), so they must be considered on both sides of the balance, i.e., pro- and contra-transparency.

Basic Transparency Zone	Pro Transparency Values	Contra Transparency Values
AI interacting with humans or generating synthetic content (e.g., deepfakes, biometric categorisation, public-interest texts).	Respect for private life Protection of personal data Freedom of expression and information Freedom of thought Human dignity Right to property	Protection of personal data Freedom of arts and sciences Right to effective remedy and fair trial Right to property Freedom of expression and information Freedom to conduct business

AI system disclosure regarding interaction or exposure to natural persons covers two cases. First, the providers must disclose AI systems directly interacting with natural persons (Article 50.1); second, the deployers must disclose the exposure to AI systems for biometric categorisation or emotion recognition (Article 50.3). In both cases, the applicability of the exception of law enforcement is subject to appropriate safeguards for the rights and freedoms of third parties. This requirement is specific for AI system disclosure on interaction or exposure for natural persons and is not needed for synthetic content disclosures.

AI systems in direct interaction with natural persons (Article 50.1)

AI disclosure duty is required when an AI system directly interacts with natural persons. Identification of AI may either be implied—when it is "obvious from the point of view of a natural person who is reasonably well-informed, observant and circumspect, taking into account the circumstances and the context of use"—or explicit, through verbal disclaimers, such as warnings, labelling, or other acknowledgements. Signals may also be behavioural (visual clues, such as bot avatar or intentionally synthetic voice).[42]

42 Michael Andrews, 'Emerging Best Practices for Disclosing AI-Generated Content' (*Kontent.ai*, 30 August 2023) <https://kontent.ai/blog/emerging-best-practices-for-disclosing-ai-generated-content/> accessed 18 December 2024.

One challenge is whether the mere awareness of AI is enough for an average person. The first presumption is that an average person, being informed about the use of AI, should be observant enough to see the notification. The second presumption is that the average person is reasonably well-informed. This means that an average person can critically assess AI content with prior and updated knowledge/experience of AI logic, risks, faults, and related issues. The third presumption is that an average person should be circumspect and have realistic expectations for AI content.

These presumptions, however, reflect the high standard of an average person. A simple generative AI chatbot interface with one line for asking questions signals to a user that the AI system is superhuman and knows everything. There is no direct duty to prove AI competence or verify the authority of the output. There is no direct duty to provide an alternative without AI automation, for example, to submit AI-generated content to a human reviewer for verification, evaluate the output's accuracy, or have mechanisms to improve the accuracy. The high standard of an average person may be closer to a "European average person," but it still is hardly compatible with the fact that, in 2024, 44% of EU citizens lack basic digital skills.[43]

The AI system disclosure duty is not applicable in cases of AI systems used for crime detection, investigation, prevention, or prosecution unless the public uses the system for crime reporting (Article 50.1). The law enforcement sector is classified as higher risk because of the power imbalance, the capability to systematically impact human rights and freedoms, and the importance of public trust, remedies, and accountability (Recital 59). However, the public interest in security and public order outweighs the public right to know about the exposure to AI; proportionality is not directly considered (though appropriate safeguards for the rights and freedoms of third parties must be applied, unless AI systems are available for the public to report a criminal offence).

The second case of AI system disclosures applies when AI is legitimately used as an element of emotion recognition or biometric categorisation (Article 50.3). For example, this includes AI-empowered detection of distractions by a driver, emotion detection from the human voice in customer support, etc. Through this, the lawmaker covers a zone beyond the average user's knowledge and recognises an impaired human capability to detect and understand AI. It also indicates a zone of heightened risk of AI misuse, bordering prohibited practices under the EU AI Act, such as real-time remote biometric identification in public places and emotion recognition for law enforcement, workplace monitoring, and education (Article 5). Unlike the first case, the duty falls on the deployer, not the provider, as the deployer has direct control over AI use. Along with personal data protection requirements, AI-powered emotion recognition or biometric categorisation is subject to the AI disclosure duty and informed choice.

43 European Union, 'Digitalisation in Europe – 2024 Edition' (*EuroStat*, 29 April 2024) <https://ec.europa.eu/eurostat/web/interactive-publications/digitalisation-2024> accessed 18 June 2024.

The standard exception of crime detection, investigation, and prevention applies to the AI used in emotion recognition and biometric categorisation; prosecution is not included. These exceptions reflect the objectively widespread practice of using biometric data in individual crime detection and investigation activities (such as forensic analysis) that could and should be more effective with the help of undisclosed AI. Considering that exceptions must be drawn and interpreted narrowly, there is a question of whether the silent imminence factor is essential when weighing values here. It seems that it is not. If it were, predictive crime prevention with undisclosed use of AI would be limited only to crimes with a high degree of objective likelihood of being committed. In this case, preventive minor financial crime screening would be impaired in efficiency due to the disclosure duty. Public order (not an exemption) and public safety (exemption, Article 3(45), 3(46), Recital 33) may overlap in this case. However, it also must be noted that the generic reason for crime prevention, not subject to disclosure duty, is very convenient to justify the excessive intervention and infringement of privacy, which may lead to factual mass surveillance. The other question is whether the duty of disclosure in crime prevention is limited to serious crimes. Real-time remote biometric identification (not categorisation) in public places is tied to limited cases of grave crimes (such as terrorism) and paired with judicial and legal oversight (Recital 35). Again, a systemic reading of the EU AI Act indicates that AI systems used for any crime prevention are exempt from the disclosure duty.

Other transparency challenges in the second case emerge in implementation, monitoring, and communication with exposed users. First, the scope of information to be disclosed and the nature of disclosure are not explicit despite the reality that AI use in borderline practices necessitates more transparency measures than in other cases. Second, it is good that the compliance burden shifts to the side of the deployer, as the user now benefits from the presumption of being a weaker side. However, the deployers are not providers; it is questionable whether they can provide enough algorithmic transparency when they do not have complete control over the knowledge of external datasets their AI systems use. Third, in some cases, transactional transparency might have flaws; for example, users might not be consciously aware that their biometric or emotional data is being collected and analysed, even if informed consent is technically given. For example, they understand the initial purpose of giving consent in a specific context but ignore the secondary uses in other contexts of the provider.

Synthetic Content Disclosures

Synthetic content disclosures cover two cases: providers' disclosure to machines that content is synthetic (technical watermarking, Article 50.3) and deployers' disclosure to persons that content is synthetic (deepfake clause and news clause, Article 50.4).

Technical *watermarking* of AI-generated or manipulated content is important for compliance, liability, risk mitigation and prevention, the quality of future AI, and recognition of digital

"waste" in the digital ecosystem. At the same time, similar to the legal challenges of counterfeit products, mechanisms to ensure authenticity and traceability are necessary.

The technical identification of AI is subject to evolving "state of the art" requirements for labelling and detection. Codes of conduct (Article 95) are expected to bring further clarity. Technical signals—via metadata, watermarking, or cryptographic signatures—must label content as AI-generated and purport information about the source and modification, if any. Technical identification signals applied to synthetic outputs of AI systems must conform to technical standards to be "effective, interoperable, robust, and reliable as far as this is technically feasible" (Article 50).

However, the current state of watermarking AI-generated text is not perfect,[44] as AI can still not recognise AI content in all cases. Additionally, watermarking may occur either during content generation or afterwards. If technical standards evolve, earlier watermarked content may not be updated for the earlier generated content.[45]

Technical watermarking exceptions cover the extent to which AI systems perform an assistive function for standard editing or do not substantially alter the input data or its semantics. A standard law enforcement exception is also applicable.

The deepfake clause (Recital 134, Article 50.4) binds deepfake deployers, explicitly requiring technical disclosure—meaning "the outputs of the AI system are marked in a machine-readable format and detectable as artificially generated or manipulated." Despite its effect on disinformation and its impact on democratic processes, deepfake is not classified as high risk by default. Deployers of deepfake have a general duty to disclose that the content has been artificially generated or manipulated, but exceptions exist.

The first exception permits the legitimate use of deepfakes for crime detection, investigation, prosecution, and prevention, exempting institutions from the disclosure duty. The second exception, derogation, still requires disclosure of exposure to AI but allows for lighter transparency requirements. If a deepfake is part of an artistic, creative, satirical, fictional, or analogous work or program, and it is evident that this work or program is such, the disclosure must be adapted to avoid hindering its display or enjoyment.

However, the normative ambiguity surrounding what qualifies as "evident" in artistic, creative, satirical, fictional, or analogous content presents implementation challenges. Logically, the concept of evident would be assessed through the prism of the "average user". However, attitudes toward satire and art may differ depending on the context,

44 Meghan Heintz, 'Watermarking for AI Text and Synthetic Proteins' (*Towards Data Science*, 7 November 2024) <https://towardsdatascience.com/watermarking-for-ai-text-and-synthetic-proteins-fighting-misinformation-and-bioterrorism-fd45be625dfe> accessed 17 November 2024.

45 Justyna Lisinska and Daniel Castro, 'The AI Act's AI Watermarking Requirement Is a Misstep in the Quest for Transparency' (*Center for Data Innovation*, 9 July 2024) <https://datainnovation.org/2024/07/the-ai-acts-ai-watermarking-requirement-is-a-misstep-in-the-quest-for-transparency/> accessed 19 July 2024.

culture, upbringing, age, and other essential circumstances. The attitude of the deployer may differ from the attitude of the public, too. As a result, multiple "average user" profiles could emerge, leading to incorrect labelling or the spread of unmarked AI-generated misinformation.

The second challenge lies in striking the balance between freedom of expression and public trust, which remains somewhat flawed. Exceptions to AI labelling or variations in AI disclosure methods may lead to situations where it is unclear what AI content is and what is not, leading to challenges in compliance, enforcement, and transparency. Highly convincing deepfakes may affect public discourse and elections, and trigger quick yet significant or irreversible changes to individual or collective decision-making. Therefore, open labelling and technical watermarking of AI-generated content should be prevalent in all deepfake cases.

Under the *synthetic news clause* (Recital 134, Article 50.4), AI can generate or manipulate text to inform the public on matters of public interest. Still, deployers must disclose that the text has been artificially generated or manipulated. This AI disclosure duty is subject to exceptions.

The first exception covers law enforcement. The second exemption applies when AI functions as an assistant to a human. In cases where AI-generated content has undergone a process of human review or editorial control and where a natural or legal person holds editorial responsibility for its publication, the disclosure requirement does not apply.

Potentially, disclosure that AI—rather than a person—created the content (for example, police briefings) would cast a shadow on the content's credibility, public trust, and acceptance. In other cases, it might compromise investigative techniques and further deepen distrust for law enforcement institutions. However, as AI detection tools improve, it remains unclear how these effects might be mitigated.

In this case, there is also no clarity on the concept of public interest and who is competent to inform the public. Typically, journalistic activities are included, but bloggers, public persons, and influencers are under question in different jurisdictions. The same applies to lawyers who may not benefit from the law enforcement exemption. There is also a slim line between assistance and predominant content creation, and it is up to humans to decide whether the line was overstepped and whether disclosure is necessary. For instance, if subscribers were to discover that up to 49% of the content in daily newspaper articles is AI-generated, they might raise doubts about the truth percentage within the texts and the fairness of pricing. Currently, only fully automated content creation is likely to be subject to labelling requirements.

These doubts also cast a shadow on the next transparency zone. The moderate transparency zone, in which general-purpose AI models are situated, presents and amplifies enforcement challenges at a scale. The next section focuses on this.

4.3. Moderate Transparency Zone

The moderate transparency zone covers free and open-source general-purpose AI (hereinafter GPAI) models, proprietary GPAI models, and GPAI models with systemic risk (Chapter V). The first attempts to translate the EU AI Act into technical requirements show that compliance level varies across large language models, with many popular ones scoring relatively low.[46] Notably, there is no duty of comparability, meaning that the quality of GPAI model as a service for downstream deployers and for business-to-customer users may be different, often favouring the provider's own users.

In the moderate transparency zone, the principle of transparency becomes interlaced with accountability, safety, and other AI principles, making it complicated to define transparency precisely or measure its changes across zones. Compared to the basic transparency zone, the moderate transparency zone adds layers of data and algorithmic transparency. It also targets a multistakeholder audience, including the public, AI provider-integrator, and supervisory authorities.

Transparency obligations include Article 53, which includes two key requirements common to all GPAI models: (1) a copyright and related rights policy and (2) a sufficiently detailed public summary of the content used for training. These requirements impose a significant administrative and technical burden on already existing GPAI models that were created without them, particularly those trained on diverse and large datasets.

For free and open-source GPAI models (Article 53.2), transparency duties are less stringent. Providers must publish key details, including parameters, weights, information on the model architecture, and model usage. In contrast, proprietary GPAIs follow stricter rules. First, GPAI model providers must provide detailed, up-to-date technical documentation on training, testing, evaluation, and capabilities. Second, they must share "acceptable use" documentation to help downstream providers understand model capabilities and limitations while complying with their obligations (Article 53).

More stringent transparency standards apply to GPAI models with systemic risk (Article 55). These models cannot benefit from the privilege of the lighter transparency requirements available to free and open-source GPAI models. Instead, they are subject to additional duties related to risk management, including model evaluation and resistance to risks, assessment and mitigation of possible systemic risks at the EU level, incident tracking, documentation, reporting and correction, and adequate level of security.

Values that participate in the balance test depend on the risk of a GPAI model. For models posing systemic risks, the impact on fundamental rights and freedoms must be assessed. In

46 Pascale Davies, 'Are AI Companies Complying with the EU AI Act? A New 'LLM Checker' Can Find Out' (*EuroNews*, 16 October 2024) <https://www.euronews.com/next/2024/10/16/are-ai-companies-complying-with-the-eu-ai-act-a-new-llm-checker-can-find-out?utm_source=substack&utm_medium=email> accessed 18 December 2024.

contrast, for models without systemic risk, the balance is rather limited, and standard tests for harm or public interest overrides are not required.

Moderate Transparency Zone	Pro Transparency Values	Contra Transparency Values
GPAI models without systemic risk or with systemic risks.	Freedom of expression Freedom of information Right to property Right to liberty and security Protection of personal data Fundamental rights and freedoms in corpore (assessments)	Right to property Freedom to conduct business Protection of personal data

GPAI models without systemic risk

The EU AI Act emphasises *proportional* transparency measures (Recital 101) for all GPAI model providers. It requires "*high* levels of transparency" for open-source GPAI models due to the blurring line of shared responsibility (Recital 102).

Open source and proprietary GPAI model transparency obligations differ in terms of their target audience and level of detail. In both cases, supervisory authorities have access to testing results and evaluation processes, but the public does not. However, the public can access the content summary for training data in both cases (Article 53.1, Recital 107). Proprietary AI systems, in contrast, must provide a more detailed version of this summary to supervisory authorities and integrators (Annex XI - XII).

Parameter disclosure, model architecture, usage instructions, and ongoing documentation are directed to the public/integrators or supervisors/integrators. Open-source AI systems publish technical documentation that is not necessarily detailed enough to disclose substantial information on training data and fine-tuning (Recital 104), and proprietary AI systems provide it confidentially for supervisory institutions and integrators. The overlap of the target audience by the public and integrators or supervisors and integrators may complicate the explainability of AI, as explainability to developers means different things than that of an average individual. Key benchmarks, such as comparison with human performance metrics, may help to understand information.[47]

47 ITI Policy Principles (n 20).

Differences in transparency for proprietary and open-source GPAIs could also lead to inconsistent compliance. Public perception of transparency often depends on publicly available information, meaning proprietary GPAIs risk being less trusted than open-source models.

GPAI providers' transparency duties towards downstream deployers (as specified in Annex XII) require them inter alia to share information with their peers that is (was) part of proprietary know-how. This disclosure often results in partial visibility and limited understandability.

First, acceptable use policies, drafted without standardised requirements, tend to reflect providers' risk aversion priorities rather than societal impact.[48] These policies often shift responsibility to deployers and may be changed unilaterally. Direct use limitations within them are challenging, especially when integrating several models.[49]

Second, indirect limitations—such as inbuilt safeguards—can subtly change model behaviour and amplify limitations to a societal risk, as seen in models like DeepSeek.

Another challenge relates to training data transparency[50]. Major AI providers opt for high-detail descriptions of training data while maintaining intentional opacity,[51] echoing the generality of the obligation to provide a copyright summary under Annex XII but renders the disclosures of little practical use. The major downside of minimal transparency is the default prioritisation of temporary competitiveness over risk management and copyright protection.

GPAI Models with Systemic Risk

The concept of systemic risk is technical. It depends on the number of parameters, users, quality and size of the dataset, amount of computation, input and output modalities, benchmarks and evaluation of capabilities, impact depth, and scope (Annex XIII of the EU AI Act). Determining whether a GPAI model poses systemic risk involves an initial self-assessment by the provider followed by regulatory authorities (Article 55 of the EU AI Act). This two-step process may create ambiguity in determining whether the risk is "systemic".

In addressing systemic risk, the EU AI Act prioritises accountability and risk management over transparency (Recital 104). However, systemic risk still affects transparency. Providers must supply more extensive and detailed documentation with a proactive

48 Kevin Klyman, 'Acceptable Use Policies for Foundation Models' (2024) 7(1) Proceedings of the AAAI / ACM Conference on AI, Ethics and Society 760, doi:10.1609/aies.v7i1.31677.

49 ibid 760.

50 ibid 753.

51 Adam Buick, 'Copyright and AI training data–transparency to the rescue?' [2024] Journal of Intellectual Property Law and Practice 3, doi:10.1093/jiplp/jpae102.

approach to risk assessment and mitigation. Information and documentation on systemic risks are subject to supervisory institutions' confidentiality obligations, meaning it is not publicly available in its entirety.

While the legislator addressed systemic risk through extra accountability, the high transparency zone offers a useful point of comparison, as it establishes transparency duties for AI systems that can also significantly impact individuals.

4.4. High Transparency Zone

High transparency is required for high-risk AI systems (Chapter III), which may account for one-fifth of all AI systems.[52] Unlike prohibited practices, high-risk AI systems are defined by their intended use.[53] In light of Article 6 of the AI Act, these systems pose a significant risk of harm to the health, safety, or fundamental rights of natural persons, including through their influence on decision-making outcomes.

The list of high-risk AI systems means slowing the innovation speed in sensitive areas to prevent quick and dangerous deterioration of Europe's fundamental values. In such domains, human oversight, decision-making, and intolerance for AI-related errors and failures outweigh AI automation's benefits by default. These areas include public security (critical infrastructure), product safety (Annex I), individual safety (emergency), privacy (profiling), law enforcement, and the administration of justice and democratic services. However, limiting access to governmental information in these contexts can lead to reduced public scrutiny.

While public access to governmental information is restricted on these grounds, high-risk AI systems operating in these areas are subject to increased supervision and confidentiality duties. Notably, the list of limitations to access information in freedom of information laws is not identical to the high-risk list in the EU AI Act. Some newly introduced areas in the EU AI Act extend protections to decision-making affecting individuals, covering areas such as education and vocational training, employment, access to services, creditworthiness, and insurance.

52 Initiative for Applied Artificial Intelligence, *AI Act: Risk Classification of AI Systems from a Practical Perspective: A study to identify uncertainties of AI users based on the risk classification of more than 100 AI systems in enterprise functions* (AppliedAI 2023) <https://www.appliedai.de/en/insights/ai-act-risk-classification-of-ai-systems-from-a-practical-perspective> accessed 18 June 2024.

53 Kazim and others (n 26) 383.

High Transparency Zone	Pro Transparency Values	Contra Transparency Values
High-risk AI systems: Biometrics, critical infrastructure, education, employment, law enforcement, asylum and border control, democratic processes, etc.	Personal data protection Respect for private life Right to an effective remedy and a fair trial Right to property Right to liberty and security Right to vote and stand as a candidate Freedom of thought, conscience and religion Right to education Freedom to choose an occupation and engage in work Consumer protection Equality before the law Right to life Environmental protection	Right to property Freedom to conduct business Freedom of the arts and sciences Protection of personal data

Derogations from the high-risk list include four cases where the risk is deemed minor: when AI is intended to perform narrow procedural tasks, improve past human work, identify decision-making patterns or deviations without influencing earlier human decisions, or assist in preparatory assessments. Recital 53 clarifies the derogations with examples for guidance that still leave ambiguity for what a "narrow" procedural task is, how to differentiate it from the preparatory task that is not required to be narrow, or what scope of content enhancement is acceptable. The profiling of natural persons in high-risk sectors is always considered high-risk. The duty of assessment and registering in the high-risk system register (Article 49) remains.

Importantly, the risk level is based on self-assessment, and businesses may be tempted to market their AI systems accordingly, even if the actual intended use falls into the high-risk zone. For example, an AI-powered search for healthcare information, where users input their symptoms, may be marketed as having basic risk. However, if the search result is an AI-generated summary with ranked potential diagnosis results and is used by professionals,

it may mean a high risk for high-stakes decision-making. The question of supervision of similar cases remains open.

Additional transparency requirements for high-risk systems are justified by factors such as the power imbalance (e.g., law enforcement), capability to have a large-scale impact on rights (Recital 66), effect international obligations (e.g., migration, Recital 60), legal decision-making (e.g., administration of justice), or democracy (e.g., elections). These transparency rules concern large-scale or high-impact risks, preventing quick decisions and radical consequences. While systemic risks are addressed by regulating GPAI models, a slow but coordinated impact— arising from multiple low-risk AI systems or targeting specific languages or regions—is not controlled by transparency rules. This regulatory gap leaves the public vulnerable to low-scale or low-impact processes.

The target audience for high-risk AI systems is at its fullest scope and includes all participants in the AI ecosystem. The largest scope of information is visible to providers, deployers and supervisory institutions, not the public. At the same time, tailored and contextual transparency measures, customised individually, may mean a more complicated approach towards removing systemic or societal risks, as they may be hidden beneath the customisation.

Providers of high-risk AI systems have eight primary obligations, set in Article 4 and 16 and Recital 72 of the AI Act. These include AI literacy, data governance, technical documentation and records, transparency, accuracy and cybersecurity, a quality management system, a declaration of conformity, and CE marking. Deployers of high-risk AI systems have four obligations under Articles 4 and 26: AI literacy, monitoring, ensuring the quality of input data, and adherence to acceptable use. Notably, deployers have noticeably more transparency obligations than those in basic and moderate transparency zones.

The provider's duty vis-à-vis downstream deployers in high-risk AI systems covers an internal dimension of transparency, aimed at equipping deployers with to fulfil their transparency obligations. The 'average deployer' concept rests on the AI provider's perception of how the provider reads and foresees deployers' needs and expertise levels.

The provider duty vis-à-vis downstream deployers (Articles 25 and 26) consists of two elements: transparency by design and transparency by instructions. First, the AI system must be designed to allow one to understand and assess how it works, its strengths and limitations, and its intended and prohibited uses—ensuring "acceptable use" by design. Second, the "acceptable use" instructions must address "possible known and foreseeable circumstances related to the use of the high-risk AI system."

To meet essential transparency requirements, three principles: practicality (obligation to provide practical examples in instructions), comprehensibility (meaningful, comprehensive, accessible, and understandable information in all provider

documentation), and language (obligation to give instructions' translations into the languages easily understandable target deployers).

A thin line separates the responsibilities of providers and deployers, with an extra layer of transparency required for public sector deployers (Article 26.8). Providers are primarily responsible for initial transparency duties, such as technical documentation, instructions of use, and registration. However, much of the responsibility is shifted to deployers, who must ensure transparency in actual usage. Deployers are accountable for input data quality, monitoring, human oversight, corrective actions, and fulfilling the right to explanation. In the workplace, they must inform employees before using high-risk AI systems. Public-sector deployers have additional transparency duties, including compliance with registration rules (Article 49), obtaining judicial authorisation, or reporting on biometric data usage (Recital 94).

In addition to the duty to conduct fundamental right impact assessments (Article 27). However, exceptions and derogations include privacy (Article 10), intellectual property rights, confidential business information, and trade secrets (Article 25), public security, a specific, substantial, and imminent threat to the life or physical safety of natural persons, the protection of life and health of persons, environmental protection, or the protection of key industrial and infrastructural assets (Article 46). AI providers and deployers are expected to demonstrate expert knowledge and impartiality when balancing these competing values—though, in practice, it does not always happen.

The hybrid transparency zone, which integrates elements from multiple transparency zones, further adds to the complexity of compliance and oversight.

4.5. Hybrid Transparency Zone

Hybrid transparency, outlined in Recital 137, arises when a particular sector has multiple transparency zones. For example, if a high-risk system is to be used in the context of specific risk under Article 50 of the EU AI Act, the transparency obligations cumulate.

Likewise, some open-source AI systems may be exempt from the EU AI Act or fall into high-risk or low-risk sectors (Article 2). AI use in elections is permitted as high risk (Annex III) and low risk. Likewise, certain law enforcement AI applications are exempt under Article 2, whereas others pose a high risk (Article 6) or prohibited risk (Article 5). Besides, a hybrid transparency zone applies when a product falls into several categories. For example, GPAI models that are customised for high-risk use may be subject to combined transparency obligations. In such cases, transparency requirements accumulate rather than being applied separately (Recital 137).

The existence of multiple transparency zones presents several risks. First, regulatory compliance and transparency measures are complicated in overlapping risk categories and become even more complex when AI systems evolve and blur category borders. For example, in agentic mesh AI, where multiple AI agents, humans and multi-step solutions

interact,[54] issues related to boundaries and ownership are likely to arise in overlapping transparency zones. Transparency is necessary to evaluate how AI agents—being more active than just tools—are using user data, communicating with other agents, and interacting with user tools.[55]

Second, the balance between transparency and secrecy differs across different risk categories, potentially creating a risk of having a "median" level of transparency that is not tailored to specific needs. Third, a loss of public trust in one category can undermine trust across all categories. Fourth, the existence of exemptions or complicated structures—such as "exceptions of exceptions"—may encourage double or varied standards of transparency. Fifth, managing these requirements demands high expertise and additional staff. Sixth, vague exceptions to rules, especially if vague, or different procedural safeguards (e.g., proportionality and harm tests) may create gaps or at least result in uneven application of transparency measures. Finally, the overlap in data sources used across various AI systems could lead to contested data quality, underscoring the need for clear transparency principles and a solid transparency framework.

5 CONCLUSION

The EU AI transparency framework is fragmented and largely dependent on the tiered risk framework. While transparency zones are coherent with predetermined risks, the differences between them are steep. More transparency benefits supervisors but not the public. The framework needs further calibration, as its numerous limitations make it too complicated to define the sensitive line between openness and confidentiality. Larger-scale risks arising from low-risk but very common AI, as well as predetermined value balance, may emerge.

One concern is the overly broad zero-transparency zone (set by Article 2), which allows three major stakeholders in the AI ecosystem—governments, the public, and businesses—to experiment with AI discreetly. AI providers enjoy broad discretion when interpreting concepts and making self-assessments, further extending the scope of zero transparency. In contrast, the basic transparency zone (set by Article 50) is too limited, creating vulnerabilities for IP rights, freedom of thought, and democratic processes. It rests on the average end user, who may not be educated enough to make an informed choice, and it overlooks the dynamic nature of AI technologies and the modest understanding of external risks.

54 Eric Broda, 'Agentic Mesh: The Future of Generative AI-Enabled Autonomous Agent Ecosystems' (*Medium*, 6 November 2024) <https://medium.com/towards-data-science/agentic-mesh-the-future-of-generative-ai-enabled-autonomous-agent-ecosystems-d6a11381c979> accessed 18 June 2024.

55 Paz Perez, 'Treating AI Agents as Personas: Introducing the Agent Computer Interaction era' (*Medium*, 5 November 2024) <https://medium.com/user-experience-design-1/treating-ai-agents-as-personas-6ef0135bdcad> accessed 18 December 2024.

The moderate transparency zone (established by Chapter V) shows other challenges, particularly regarding fluctuating responsibility between providers and downstream deployers. The contrast in rules for open-source and proprietary software demonstrates an overestimation of the ability and willingness to manage the inbuilt risks of open-source software. At the same time, the high transparency zone (established by Chapter III) disproportionately privileges law enforcement. Disclosure obligations are not significantly increased, but the scope of risk management is, shifting the burden onto supervisory authorities.

A critical issue emerges in the hybrid transparency zone, where interactions between different AI systems are not well addressed. With the rise of agentic AI and the further evolution of AI technologies, these interactions may materialise in unpredictable and unforeseen ways.

Despite these flaws, the EU AI Act is a progressive and ambitious attempt to create a human-centered AI environment. Its long-term success, inter alia, depends on solidifying, simplifying and strengthening the AI transparency framework. Systemic improvements include establishing clear disclosure principles and milestones, empowering the public and media with participatory rights, strengthening AI literacy, establishing dogmatic conflict resolution rules, and reframing transparency to have more impact on confidentiality. Without them, AI transparency risks being a multifaceted compromise rather than a guiding principle that helps achieve the balance promised in the EU AI Act.

Future research on transparency in AI-powered digital services, small language models, governmental AI usage, the transparency obligations of providers, or the public role may help identify more underexplored gaps in AI regulation. Addressing these challenges proactively is important for a sustainable AI governance framework and a better future.

REFERENCES

1. Andrews M, 'Emerging Best Practices for Disclosing AI-Generated Content' (*Kontent.ai*, 30 August 2023) <https://kontent.ai/blog/emerging-best-practices-for-disclosing-ai-generated-content/> accessed 18 December 2024.
2. Broda E, 'Agentic Mesh: The Future of Generative AI-Enabled Autonomous Agent Ecosystems' (*Medium*, 6 November 2024) <https://medium.com/towards-data-science/agentic-mesh-the-future-of-generative-ai-enabled-autonomous-agent-ecosystems-d6a11381c979> accessed 18 June 2024.
3. Buick A, 'Copyright and AI training data–transparency to the rescue?' [2024] Journal of Intellectual Property Law and Practice jpae102, doi:10.1093/jiplp/jpae102.
4. Colonna L, 'The AI Act's Research Exemption: A Mechanism for Regulatory Arbitrage?' in Gill-Pedro E and Moberg A (eds), *YSEC Yearbook of Socio-Economic Constitutions 2023: Law and the Governance of Artificial Intelligence* (Springer Cham 2023) 51, doi:10.1007/16495_2023_59.

5. Davies P, 'Are AI Companies Complying with the EU AI Act? A New 'LLM Checker' Can Find Out' (*EuroNews*, 16 October 2024) <https://www.euronews.com/next/2024/10/16/are-ai-companies-complying-with-the-eu-ai-act-a-new-llm-checker-can-find-out?utm_source=substack&utm_medium=email> accessed 18 December 2024.
6. Draghi M, *The Future of European Competetiveness: Part A: A Competiteveness Strategy for Europe* (European Commission 2024).
7. Ebers M, 'Truly Risk-Based Regulation of Artificial Intelligence: How to Implement the EU's AI Act' [2024] European Journal of Risk Regulation 1, doi:10.1017/err.2024.78.
8. Fanni R, 'Why the EU Must Now Tackle the Risks Posed by Military AI' (*CEPS*, 8 June 2023) <https://www.ceps.eu/why-the-eu-must-now-tackle-the-risks-posed-by-military-ai/> accessed 18 December 2024.
9. Fraser H and Bello y Villarino JM, 'Acceptable Risks in Europe's Proposed AI Act: Reasonableness and Other Principles for Deciding How Much Risk Management Is Enough' (2024) 15(2*)* European Journal of Risk Regulation 431, doi:10.1017/err.2023.57.
10. Frattone C, 'Reasonable AI and Other Creatures: What Role for AI Standards in Liability Litigation?' (2022) 1(3) Journal of Law, Market & Innovation 15, doi:10.13135/2785-7867/7166.
11. Geng Y, 'Transparency for What Purpose?: Designing Outcomes-Focused Transparency Tactics for Digital Platforms' (2024) 16(1) Policy & Internet 83, doi:10.1002/poi3.362.
12. Haataja M and Bryson JJ, 'What Costs Should We Expect from the EU's AI Act?' (*Center for Open Science*, 27 August 2021) SocArXiv 8nzb4, doi:10.31235/osf.io/8nzb4.
13. Haresamudram K, Larsson S and Heintz F, 'Three levels of AI transparency' (2023) 56(2) Computer 93, doi:10.1109/MC.2022.3213181.
14. Hauer MP and others, 'Quantitative Study About the Estimated Impact of the AI Act' (*arXivLabs*, 29 March 2023) arXiv:2304.06503 [cs.CY], doi:10.48550/arXiv.2304.06503.
15. Heintz M, 'Watermarking for AI Text and Synthetic Proteins' (*Towards Data Science*, 7 November 2024) <https://towardsdatascience.com/watermarking-for-ai-text-and-synthetic-proteins-fighting-misinformation-and-bioterrorism-fd45be625dfe> accessed 17 November 2024.
16. Holst L and others, 'The Impact of the EU AI Act's Transparency Requirements on AI Innovation' (19th International Conference on Wirtschaftsinformatik (WI), Würzburg, Germany, September 2024) <https://eref.uni-bayreuth.de/id/eprint/90313/> accessed 17 November 2024.
17. Hosain MT and others, 'Path to Gain Functional Transparency in Artificial Intelligence with Meaningful Explainability' (2023) 3(2) Journal of Metaverse 166, doi:10.57019/jmv.1306685.

18. Jedličková A, 'Ethical Considerations in Risk Management of Autonomous and Intelligent Systems' (2024) 14(1-2) Ethics & Bioethics 80, doi:10.2478/ebce-2024-0007.

19. Kazim E and others, 'Proposed EU AI Act–Presidency Compromise Text: Select Overview and Comment on the Changes to the Proposed Regulation' (2023) 3 AI Ethics 381, doi:10.1007/s43681-022-00179-z.

20. Kelly J and others, 'Navigating the EU AI Act: A Methodological Approach to Compliance for Safety-Critical Products' (*arXivLabs*, 26 March 2024) arXiv:2403.16808 [cs.AI], doi:10.1109/CAI59869.2024.00179.

21. Khan T and Srikumar M, 'Developing General Purpose AI Guidelines: What the EU Can Learn from PAI's Model Deployment Guidance' (*Partnership on AI*, 26 November 2024) <https://partnershiponai.org/developing-general-purpose-ai-guidelines-what-the-eu-can-learn-from-pais-model-deployment-guidance/> accessed 18 December 2024.

22. Kieseberg P and others, 'Controllable AI - An Alternative to Trustworthiness in Complex AI Systems?' in Holzinger A and others, (eds), *Machine Learning and Knowledge Extraction: 7th IFIP TC 5, TC 12, WG 8.4, WG 8.9, WG 12.9 International Cross-Domain Conference, CD-MAKE 2023, Benevento, Italy, 29 August – 1 September 2023* (LNCS 14065, Springer Cham 2023) 1, doi:.1007/978-3-031-40837-3_1.

23. Kiseleva A, Kotzinos D and De Hert P, 'Transparency of AI in healthcare as a multilayered system of accountabilities: between legal requirements and technical limitations' (2022) 30(5) Frontiers in artificial intelligence 879603, doi:10.3389/frai.2022.879603.

24. Klyman K, 'Acceptable Use Policies for Foundation Models' (2024) 7(1) Proceedings of the AAAI / ACM Conference on AI, Ethics and Society 752, doi:10.1609/aies.v7i1.31677.

25. Lisinska J and Castro D, 'The AI Act's AI Watermarking Requirement Is a Misstep in the Quest for Transparency' (*Center for Data Innovation*, 9 July 2024) <https://datainnovation.org/2024/07/the-ai-acts-ai-watermarking-requirement-is-a-misstep-in-the-quest-for-transparency/> accessed 19 July 2024.

26. Nannini L, 'Habemus a Right to an Explanation: so What? – A Framework on Transparency-Explainability Functionality and Tensions in the EU AI Act' (2024) 7(1) Proceedings of the AAAI / ACM Conference on AI, Ethics, and Society 1023, doi:10.1609/aies.v7i1.31700.

27. Novelli C and others, '*AI Risk Assessment: A Scenario-Based, Proportional Methodology for the AI Act*' (2023) 3 Digital Society 13, doi:10.1007/s44206-024-00095-1.

28. Perez P, 'Treating AI Agents as Personas: Introducing the Agent Computer Interaction era' (*Medium*, 5 November 2024) <https://medium.com/user-experience-design-1/treating-ai-agents-as-personas-6ef0135bdcad> accessed 18 December 2024.

29. Powell R, 'The EU AI Act: National Security Implications' (*CETaS Explainers*, 31 July 2024) <https://cetas.turing.ac.uk/publications/eu-ai-act-national-security-implications> accessed 18 December 2024.

30. Prifti K and others, 'From Bilateral to Ecosystemic Transparency: Aligning GDPR's Transparency Obligations with the European Digital Ecosystem of Trust' in Kuhlmann S and others (eds), *Transparency or Opacity: A Legal Analysis of the Organization of Information in the Digital World* (Nomos 2023) 115, doi:10.5771/9783748936060-115.

31. Siegmann C and Anderljung M, *The Brussels Effect and Artificial Intelligence: How EU Regulation Will Impact the Global AI Market* (Center for the Governance of AI 2022).

32. Sloane M and others, 'Introducing Contextual Transparency for Automated Decision Systems' (2023) 5 Nature Machine Intelligence 187, doi:10.1038/s42256-023-00623-7.

33. Varošanec I, 'On the Path to the Future: Mapping the Notion of Transparency in the EU Regulatory Framework for AI' (2022) 36(2) International Review of Law, Computers & Technology 95, doi:10.1080/13600869.2022.2060471.

AUTHORS INFORMATION

Gintare Makauskaite-Samuole
Institute of International and EU Law, Mykolas Romeris University, Lithuania
gintare.samuole@gmail.com
https://orcid.org/0009-0007-7777-1873
Corresponding author, responsible for writing, reviewing and editing this article.

Competing interests: No competing interests were disclosed.

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

ACKNOWLEDGEMENTS

This research is part of the project The Jean Monnet Center of Excellence 'European Fundamental Values in Digital Era', 101085385 – EFVDE – ERASMUS-JMO-2022-HEI-TCH-RSCH. Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or EACEA. Neither the European Union nor the granting authority can be held responsible for them.

ABOUT THIS ARTICLE

Cite this article
Makauskaite-Samuole G, 'Transparency in the Labyrinths of the EU AI Act: Smart or Disbalanced?' (2025) 8(2) Access to Justice in Eastern Europe 38-68 <https://doi.org/10.33327/AJEE-18-8.2-a000105>

DOI https://doi.org/10.33327/AJEE-18-8.2-a000105

Managing editor – Mag. Bohdana Zahrebelna. **Section Editor** – Prof. Tetiana Tsuvina.
English Editor – Julie Bold. **Ukrainian language Editor** – Lilia Hartman.

Keywords: *AI transparency, EU AI Act, transparency zones, regulatory limitations, innovation vs. transparency.*

DETAILS FOR PUBLICATION

Date of submission - 22 Dec 2024
Date of acceptance - 06 Feb 2025
Date of Online First publication: 26 Mar 2025
Last Publication: 14 May 2025
Whether the manuscript was fast tracked – No
Number of reviewer report submitted in first round – Three
Number of revision rounds – 1 round, revised version submitted 06 Feb 2025

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

RIGHTS AND PERMISSIONS

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

АНОТАЦІЯ УКРАЇНСЬКОЮ МОВОЮ

Дослідницька стаття

ПРОЗОРІСТЬ У ЛАБІРИНТАХ ЗАКОНУ ЄС ПРО ШІ: РОЗУМНО ЧИ НЕЗБАЛАНСОВАНО?

Гінтаре Макаускайте-Самуоле

АНОТАЦІЯ

Вступ: *Повної прозорості у сфері штучного інтелекту досягти неможливо.*[1] *У взаємозалежному технологічному контексті обсяг прозорості штучного інтелекту та логіка цінностей, які переважають над прозорістю, залишаються незрозумілими. Законодавство про штучний інтелект, наприклад, Закон Європейського Союзу про штучний інтелект (далі - Закон ЄС про ШІ), намагається визначити справжнє значення і роль прозорості ШІ.*

Методи: *Автор застосовує доктринальне дослідження та методи порівняльного аналізу для оцінки прозорості штучного інтелекту в Законі ЄС про ШІ; встановлюються межі щодо чітких зон прозорості. Доктринальне дослідження допомагає визначити обсяг зобов'язань щодо прозорості та вивчити їхні обмеження та взаємодію в Законі ЄС про штучний інтелект, тоді як порівняльний аналіз висвітлює неузгодженості, такі як непояснена різниця між обов'язками щодо прозорості в окремих зонах або різні вимоги до відкритого та пропрієтарного ШІ.*

Результати та висновки: *Результати дослідження розкривають фрагментарну та нерівномірну структуру прозорості штучного інтелекту в Законі ЄС про ШІ, сформовану багатьма винятками, відступами, лімітами та іншими обмеженнями. Зона нульової прозорості (встановлена статтею 2) є надто широкою і надає значну свободу дій зацікавленим сторонам. На противагу цьому, основна зона прозорості (встановлена статтею 50) є надто вузькою, що створює ризики для фундаментальних прав людини. Наступна зона, зона помірної прозорості (Розділ V), має проблеми з розподілом відповідальності між постачальниками штучного інтелекту та тими, хто розгортає технології. Тим часом зона високої прозорості (описана в Розділі III) надає перевагу правоохоронним органам. Нарешті, гібридна зона прозорості підкреслює труднощі в управлінні взаємодією між системами ШІ з різними рівнями ризику.*

Автор робить висновок, що Закон ЄС про штучний інтелект є прогресивним, але потребує доопрацювання, щоб функціонувати як цілісна та надійна система прозорості. Шкала між суспільним інтересом до прозорості штучного інтелекту, індивідуальними та суспільними правами та законними інтересами ризикує бути відкаліброваною постфактум.

Ключові слова: *прозорість ШІ, Закон ЄС про штучний інтелект, зони прозорості, регуляторні обмеження, інновації проти прозорості.*

1 Mona Sloane and others, 'Introducing Contextual Transparency for Automated Decision Systems' (2023) 5 *Nature Machine Intelligence 188*, doi:10.1038/s42256-023-00623-7.