Article

DOI: 10.17803/2713-0533.2025.4.34.801-828

The Principle of Transparency in AI Regulation: A Comparative Legal Analysis of the European and Chinese Approaches

Yulia Kharitonova,[1] **Gergely Ferenc Lendvai**[2]

[1] *Lomonosov Moscow State University, Moscow, Russian Federation*

[2] *Ludovika University of Public Service, Budapest, Hungary*

Corresponding Author — Yulia Kharitonova

Abstract: The paper examines the principle of transparency in artificial intelligence (AI) regulations in two legal frameworks, namely, the European Union (EU) and China. The study aims to explore how transparency, a key principle for ensuring accountability and fostering trust in AI technologies, is regulated in these two distinct geopolitical environments. Using a comparative legal analysis approach, the paper reviews primary legal documents, scholarly literature, and expert analyses to identify commonalities and divergences in AI transparency regulations. The findings indicate that the EU's AI Act emphasizes a risk-based approach, categorizing AI systems into high-risk, limited-risk, and minimal-risk categories, with stringent transparency requirements for high-risk systems. These requirements include comprehensive documentation, human oversight, and explainability to ensure that AI systems operate within ethical and legal boundaries. However, the AI Act also holds challenges, particularly for smaller enterprises, in meeting these transparency demands, as well as the technical difficulties in achieving transparency in complex AI models. In contrast, China's regulatory framework, while similarly focused on transparency, integrates socialist moral and ethical values. The Chinese approach categorizes AI systems based on risk and emphasizes the interpretability and explainability of AI systems to ensure compliance with state-sanctioned

moral principles. The findings suggest that while both the EU and China recognize the importance of transparency, their regulatory frameworks reflect broader cultural and political differences. The study concludes that achieving harmonized global AI transparency standards will require ongoing technological innovation, legal refinement, and international cooperation.

Keywords: Artificial intelligence; AI Act; European law; Chinese Law; transparency

Acknowledgments: TKP2021-NKTA-51 has been implemented with the support provided by the Ministry of Culture and Innovation of Hungary from the National Research, Development and Innovation Fund, financed under the TKP2021-NKTA funding scheme. The article was supported by the ADVANCED_24 funding scheme of the Ministry of Culture and Innovation of Hungary from the National Research, Development and Innovation Fund.

Cite as: Kharitonova, Yu. and Lendvai, G.F., (2025). The Principle of Transparency in AI Regulation: A Comparative Legal Analysis of the European and Chinese Approaches. *Kutafin Law Review*, 12(4), pp. 801–828, doi: 10.17803/2713-0533.2025.4.34.801-828

Contents

I. Introduction

Artificial intelligence (hereinafter AI) dominates contemporary technological advancement: it sectors both in and outside the technological industry, amends the philosophy of governance and policy-making and steps up as one of the leading topics in scholarly work too. It can no longer be argued that AI is merely a matter of the future. From AI generating systems such as ChatGPT to algorithmic processes and deepfake technologies, as well as a wide range of other applications including autonomous vehicles and automated decision-making, AI has become an integrated and undeniably fundamental component of daily life. The law is mandated with, among other duties, to follow up on such advancements with regulatory efforts. As manifested around the world, different approaches emerge: the liberal framework set forth by the United States, the comprehensive, holistic regulatory aspect proposed by the European Union or the more rigorous legislation on such system in Russia or China might be key examples to support this statement. In this paper a significant principle will be discussed, namely, the legislative principle of transparency in AI. The choice of transparency to be covered is two-fold.

First, transparency emerged as a pivotal principle to ensure accountability, mitigate risks, and foster trust in AI technologies. Second, AI transparency within a global legal context is worth exploring in a comparative review as the principle itself stems from its profound implications for society, governance, and innovation not just in local segments but also in a global perspective. Transparency in this context is overarching in terms of scope: it concerns deployment of AI in certain regions and sectors, bias in systems and even raises critical issues in data governance (and privacy) as well as AI ethics, philosophy of technology and — as presented in this paper — regulation.

It is imperative to introduce the concept of "transparency" first. Literature on transparency in AI is highly interdisciplinary ranging from computer sciences to ethics and philosophy. The definition of transparency is, however, rather complex as each of these disciplines interpret it in a different way. In general, transparency — as a prerequisite of trust in society per Larsson and Heintz (2020) — in AI regulation is

a multidisciplinary concept which entails ensuring that AI systems are open and explainable, understandable, and accountable to their users, operators, and affected parties (Arrieta et al., 2019; Walmsley, 2020; Raja and Zhou, 2023). Larsson and Heintz (2020) also highlight the normative implications of transparency, suggesting that it often carries positive connotations linked to knowledge and trustworthiness, which influence regulatory debates.

In AI regulation transparency serves multiple purposes. On the one hand, it is a principle that is guided by the safeguarding of users and fundamental rights — key legislative imperatives that are almost omnipresent in the newer Internet and technology regulations (Söderlund et al., 2024). On the other hand, transparency also serves as a possible solution or mitigator to the black box problem (Diakopoulos and Koliska, 2017; Busuioc et al., 2022), allowing for a better understanding of how AI-powered, autonomous systems work (Papadouli, 2022).

Naturally, the detailed problematization of black boxes, especially the lack of regulation thereof, would fall out of the scope of the paper; however, it is important to outline the issue briefly. Generally, black boxes are understood as parts or elements of systems, devices, or models whose internal mechanisms are hidden or opaque to the observer (Miller, 2018; Castelvecchi, 2016). As Durán and Jongsma (2021) underline, the key problems with black boxes, from an epistemological standpoint, are the many concerns their existence raises, from the lack of transparency to potential biases that are embedded in the systems without the ability to directly assess them.

To contextualize black boxes in the field of AI regulation, black boxes are a critical legislative issue, and it can be said that regulations pertaining to AI are advocating for more explainability and transparency in AI systems for the very reason to eliminate black boxes. Nonetheless, the effectivity of regulations is questionable. Despite sector-specific regulations (Mourby et al., 2021), as there is “no easy solution” to demand explainability using the legal stipulations due the inherent (or rather, innate) nature of black boxes in AI systems, the lack of legal tools, and naturally, the intrinsic nature of AI systems (Gryz and Rojszczak, 2021).

In this regard, we should mention that transparency is extremely hard to achieve on an adequate level. In particular, with respect to more complex AI models and large language models (LLMs), research has highlighted significant technical challenges that frequently impede the effective realization of transparency (Walmsley, 2020) and there is also a heated debate in how transparency can be attained without breaching already existing regulations on data accuracy and privacy (Chaudhary, 2024; Hosain et al., 2023).

Our paper aims to conduct a comparative legal analysis of transparency regulations in AI across two major geopolitical entities: the EU and China. An important limitation should be underlined here. As noted above, while several jurisdictions, including the United States, have developed noteworthy approaches to AI regulation (albeit with indications of a downward shift in regulatory momentum under the current Trump administration (Novelli et al., 2025)), this study intentionally confines its analysis to the two regions/countries previously discussed. The reason for this selection lies in the fact that these two subjects are seemingly leading the "race" of AI regulation. In contrast, the United States' AI governance remains highly sectoral, fragmented, and largely reliant on non-binding guidelines rather than enforceable federal legislation (Litwin and Racabi, 2024), making a three-fold comparison highly difficult.

In researching the EU's and China's approaches, we aim to highlight a comparative, analytical perspective of existing frameworks. Scholarship has been growing exponentially on specific aspects of regulation (cf. Maslova and Sorokova, 2022; Vesnic-Alujevic et al., 2020); however, there is an important research gap to be filled that includes a comparative assessment of two of the most prominent regulatory frameworks. By examining the development, goals, transparency stipulations, and critical points of each regulatory framework, the research seeks to identify commonalities, divergences, and potential implications for global AI governance.

To conduct this research, the study uses a comparative legal analysis approach to review primary legal documents, scholarly literature, and expert analyses pertaining to AI regulations in the regions examined. This methodological framework allows for a nuanced exploration of how

different jurisdictions approach transparency in AI governance, thereby informing discussions on potential legal convergence or divergence. The study contributes to the existing literature by offering a comprehensive analysis of transparency as a regulatory tool in the context of digital technologies. Lastly, the research also proposes an inclusive approach to AI scholarship that invites experts and researchers from the BRICS countries to conduct research in a comparative approach.

II. The AI Act: Examination of the European Legislation

The Artificial Intelligence Act[1] (AIA) represents a significant milestone in the regulation of artificial intelligence, marking a substantive breakthrough in the establishment of a comprehensive legal framework for its governance (Helberger and Diakopoulos, 2022). A pioneering proposal by the European Union, the AI Act seeks to introduce the first comprehensive international regulation on AI systems presenting novelties such as the risk-based approach to AI and the accentuation of safeguarding fundamental rights, as well as constitutional values in an ecosphere where AI is ever-present (Tartaro, 2023).

The idea of a European artificial intelligence regulation began in 2018 when the European Commission unveiled its AI strategy in the Coordinated Plan on Artificial Intelligence (CPAI) document, a joint proposal between EU Member States, the Commission, Norway and Switzerland. The CPAI was ambitious in achieving a number of significant goals; it emphasized the need for a robust and human-centric approach to AI, drew up a plan to better support European AI infrastructure and urged the acceleration of investments in AI technologies and the AI-sector development. This initiative was followed by extensive consultations with stakeholders, including industry experts, academics, and civil society organizations. It should also be mentioned that a journalist leaked the consolidated text of the Artificial Intelligence Act

[1] Regulation (EU) 2024/1689 of the European Parliament and of the Council dated 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.

(AIA) in early 2024; however, the final, authoritative version of the Act was officially published in the Official Journal of the European Union on 13 June 2024.

II.1. Objectives and Risks

The AIA is designed to achieve a plethora of objectives. First, it seeks to ensure that AI systems used within the EU are safe and respect fundamental rights as mentioned above. Though seemingly evident, this approach is critical as the EU aims to widen the scope from a strictly legal regulation to a legal-ethical one; one which regards the ethical and constitutional values of developing a human-friendly AI environment as a paramount goal. Second, it aims to ensure that the Act promotes trust in AI technologies by implementing a risk-based approach to regulation, categorizing AI systems based on their potential impact on individuals and society (Tartaro, 2023). Artificial Intelligence systems (AIS), as defined in Art. 3(1) of the AIA, are machine-based systems designed to function with varying degrees of autonomy, capable of potential adaptation following deployment, and processing inputs to generate outputs such as predictions, content, recommendations, or decisions that affect physical or virtual environments. These systems are classified into three categories, namely, high-risk, limited-risk, and minimal-risk, with high-risk systems being subject to the most stringent regulatory requirements (Novelli et al., 2023). The differentiation between risks is as follows:

1. High-risk AIS: typically, systems used in critical infrastructure, healthcare, and law enforcement that must comply with rigorous safety, risk management, and continuous monitoring requirements due to their significant impact on safety and fundamental rights. One such example is the need for appropriate documentation and high level of compliance with existing standards (Golpayegani et al., 2023).

2. Limited-risk AIS: AIS that are subjected to moderate regulations requiring safeguards to prevent misuse and ensure transparency, addressing potential impacts without the rigid measures applied to high-risk systems.

3. Minimal-risk AIS: AIS that pose negligible risks. They are largely exempt from regulatory oversight but must still adhere to basic transparency principles to inform users about AI interactions.

It is essential to highlight the category of prohibited AI systems, as set forth in Art. 5 of the AIA, which serve as a critical mechanism for safeguarding users. Certain AI systems are banned due to their potential to cause significant harm, including those that employ subliminal techniques to manipulate behavior, exploit the vulnerabilities of specific demographic groups, implement social scoring by public authorities, or utilize real-time remote biometric identification in publicly accessible spaces for law enforcement purposes, except under narrowly defined and strictly regulated circumstances. A critical point here is that such AIS shall not be developed, placed on the market, or used within the EU, meaning that the AIA is narrowing down the scope of possible AI development and business in the Union drastically. Lastly, it is to be underlined that the AIA aims to promote European innovation in the AI sector by creating a "well-functioning internal market" for both AIS and other AI technologies (Pošćić and Martinović, 2022).

II.2. Transparency as a Principle and Obligation

Transparency is a fundamental principle embedded within the AIA to ensure the accountability and ethical deployment of AIS. The Act outlines specific obligations for AI system providers and deployers to enhance transparency and ensure that users and stakeholders are well-informed about AI operations, their capabilities, and limitations. In this context, Art. 4a serves as a cornerstone provision for ensuring transparency, establishing that "AI systems shall be developed and used in a way that allows appropriate traceability and explainability while making humans aware that they communicate or interact with an AI system as well as duly informing users of the capabilities and limitations of that AI system and affected persons about their rights."

As mentioned above, high-risk AIS face strict obligation that also encompasses a comprehensive set of transparency stipulations. Article 13 of the AIA requires high-risk AIS to be designed and developed with a high level of transparency to enable deployers to

understand and appropriately use the system's outputs. The stipulation specifies that such AISs must be accompanied by clear, comprehensive, and easily accessible instructions. These instructions should include detailed information on the system's characteristics, functionalities, and limitations, as well as the specific context in which the system is intended to operate. Furthermore, extensive documentations must be provided to give insights into the system's accuracy, robustness, and cybersecurity measures. The documentation should also outline any known or foreseeable risks related to health, safety, and fundamental rights. The objective is to equip deployers with all necessary information to make informed decisions and use the AIS effectively and safely.

Table 1: Transparency obligations for high-risk AIS

Area	**Description**
System Characteristics	Detailed description of the AI system's design, purpose, and tasks it can perform
Capabilities	Information on the system's abilities, including data processing capacity and decision-making processes
Limitations	Explanation of the boundaries within which the system operates effectively, and conditions for performance degradation
Context of Use	Specific scenarios and environments where the AI system is intended to be deployed, including prerequisites for operation
Accuracy	Data on the system's performance accuracy, including error rates and confidence levels
Robustness	Measures to ensure the system's reliability and consistent performance under various conditions
Cybersecurity Measures	Protocols and safeguards to protect the AI system from cyber threats and ensure data integrity
Risk Assessment	Identification of potential risks to health, safety, and fundamental rights, along with mitigation strategies

Moreover, Art. 12 mandates comprehensive logging and record keeping for high-risk AISs to ensure transparency, traceability, and accountability. Providers must maintain automatic logs of relevant events, document system characteristics, manage data, and record system performance and updates. Human oversight measures and

incident reports must also be documented. These records facilitate compliance verification, support incident management, and enhance trust in AI systems. Access to these records is required for audits and inspections, ensuring providers can be held accountable. Furthermore, human oversight and explainability are also crucial components ensuring that AISs operate within ethical and legal boundaries. Providers of high-risk AI systems are required to implement measures that enable human operators to understand, interpret, and control the AI system's outputs (Art. 14). This includes designing systems with technical features that facilitate human intervention when necessary, ensuring that decisions influenced by AI can be reviewed and challenged. Explainability involves providing clear, detailed information about how the AI system processes data and generates outputs, making it possible for human operators to understand the reasoning behind AI decisions (Art. 13). This is crucial for maintaining accountability and preventing unintended consequences by ensuring humans remain in control of AI systems (Art. 14).

Transparency requirements, nonetheless, also extend to AI systems that do not fall within the high-risk category. Article 50 outlines the transparency obligations for providers and deployers of certain AIS, accentuating the need for clear disclosure when individuals interact with AI systems. Providers must ensure that AI systems intended to interact directly with natural persons inform them that they are interacting with an AI system, unless it is obvious to a reasonably well-informed person. Furthermore, AI systems generating synthetic content, such as a text, audio, image, or video, must mark their outputs as artificially generated or manipulated. This requirement is intended to prevent deception and to ensure that users are fully aware of the artificial nature of the content with which they interact.

An important aspect to highlight in this context is the set of rules governing the disclosure of synthetic content. For AI systems that generate or manipulate content, such as text, audio, images, or videos (commonly referred to as deepfakes, though the AIA's definition is quite vague in this regard), Art. 50 mandates clear labeling of such outputs. Providers must mark these outputs as artificially generated or manipulated to prevent the spread of misinformation and ensure that users can distinguish between real and synthetic content. This requirement is particularly

important in contexts where the artificial content might be perceived as genuine, potentially leading to misinformation or deception.

Table 2: Comparison between transparency stipulations in the AIA

Obligations	High-Risk AI Systems	Other AI Systems
Documentation and Instructions	Must include detailed, clear, and accessible instructions covering system characteristics, capabilities, limitations, and context of use	Not explicitly required, but basic transparency and accountability principles must be adhered to
Disclosure of AI Interaction	Must inform users that they are interacting with an AI system	Required when interaction with natural persons is not obvious
Synthetic Content Disclosure	Outputs like text, audio, images, or videos must be clearly marked as artificially generated or manipulated. (Only if applicable)	Generally required to prevent deception, but specific labeling is less stringent than high-risk systems
Logging and Record-Keeping	Must include automatic logging capabilities to record relevant events, data, and interactions	Not required, but providers should maintain basic records for accountability
Human Oversight and Explainability	Must implement measures for human oversight and ensure outputs can be interpreted and controlled by humans	Encouraged to have human oversight, but not as rigorously mandated as for high-risk systems

II.3. Too little or too much? A Critique of Transparency in the context of Compliance and Deepfakes

The AIA is undoubtedly a remarkable first step in international AI regulation. Alas its comprehensive nature, the Regulation is not without flaws. In this segment, a few key issues and challenges will be identified, with particular attention to interpretation and deepfake content.

As detailed above, high-risk AIS face a multitude of transparency obligations. Although the legislator's objective is clear and commendable, the comprehensive nature of the documentation required for high-risk

AI systems pose notable challenges. As seen from the comparative table above, this level of detail necessitates substantial resources and expertise, which can be particularly burdensome for small and medium-sized enterprises (SMEs) and startups that may lack the necessary infrastructure and financial capability to meet these demands.

On the other hand, the requirement for continuous updates and record keeping, as outlined in Art. 12, adds another nuance of complexity to the already extremely complicated compliance issue. In this regard, high-risk AI systems must maintain logs and records that document their performance, including data regarding system failures and instances of non-compliance creating an ongoing obligation procedure that necessitates tiresome and vigorous data management of both systems and processes.

With regard to compliance and interpretation, the principle of explainability must be emphasized as a fundamental requirement. Article 14 requires that these systems provide clear and understandable explanations of their decision-making processes and outcomes. While this is unarguably critical for accountability and user trust, achieving explainability in complex AI models, such as deep learning networks, is inherently challenging due to their opaque nature. This challenge is further exacerbated when AI systems are required to function in real time, as the provision of timely explanations becomes both critical for effective oversight and particularly demanding from a technical standpoint.

An even more pressing issue is the insufficient regulation of emerging technologies such as deepfakes (Moreno, 2024). As argued notoriously by Gosztonyi and Lendvai (2023; 2024) among other scholars (Birrer and Just, 2024), deepfakes present a unique and problematic aspect of synthetic media regulation. Research has covered that deepfakes can be and is mostly used maliciously to spread unconsented porn (either revenge porn or so-called deepnudes) (Mania, 2022), political disinformation and to commit financial fraud, identity theft or violation of privacy.

The Act addresses these issues by imposing specific transparency requirements on AI systems that generate deepfakes, necessitating clear disclosure of their synthetic nature. However, identifying and

regulating deepfakes is technically challenging due to their increasingly sophisticated nature. Ensuring compliance with transparency rules in this context requires advanced detection technologies and continuous monitoring, which can be resource-intensive and it still not guarantees that the polemic trends in deepfake-making changes.

To put it into a practical example, it seems highly unlikely that someone who is developing deepfake porn or generates political deepfake propaganda will meticulously follow the AIA's stipulation to set out a clear disclosure that the given pornographic content or political disinformation is indeed a deepfake content. One can also argue that the above polemic lies in the interconnected nature of deepfake proliferation, and the limitations of regulatory frameworks designed primarily around transparency. The Act's reliance on disclosure *assumes* that actors producing deepfakes will comply with regulatory mandates — a presumption which, more often than not, is at odds with the harmful intent driving their creation.

Furthermore, as deepfake technology advances through ongoing technological innovation and the increasing democratization of its tools, the techniques employed to evade detection and exploit gaps in regulatory oversight likewise become more sophisticated. This dynamic creates a legislative "cat-and-mouse game" where regulatory responses must continually adapt to keep pace with technological advancements and evolving tactics in deepfake production. Therefore, the law in itself will never be enough; without bolstering detection capabilities, enforcing proactive regulations, and fostering international cooperation, the AIA risks being outpaced by the relentless ingenuity of those exploiting deepfakes.

III. Lessons from Chinese Law

Though scholarship has demonstrably covered the AIA, research on other AI regulations is rather scarce. In this context, our study examines the Chinese AI regulation, a legislative approach which aims at respecting socialist morality and ethics in the application of AI in the spirit of "developing science and technology for good" and "human-centered approach." Chinese researchers emphasize that the application

of AI not only increases efficiency in certain areas of human activity, but also contributes to the development of the economy and society (Yao and Li, 2023). Some researchers argue that the impact of technology on human society is profound, given the idea that "technology development leads to social evolution" (Pu and Xiang, 2023). At the same time, scholars take the cautious position that the risks associated with AI are difficult to predict in the early stages of technological development, although at the current stage the ability to control technological development is relatively strong. However, as technology evolves, our ability to control technology will become extremely limited because at that time "technology will have acquired sufficient power and its own path of development" (Shen, 2024, p. 73).

Security, privacy and discrimination form the basis of AI regulation in China. In the development and use of AI, relevant actors may introduce algorithmic bias and Big data bias into AI in the process of task construction, data analysis and selection of performance criteria, leading to discriminatory outcomes. These tasks determine the focus of specific legal provisions. The principle of AI transparency is considered in Chinese law in direct relation to AI explainability and interpretability, which is referred to as the link between AI and law. Given that "only what is already understandable can be legally formalized," the interpretability of AI has become a necessary condition for the promotion and application of AI, as well as for addressing its legal liability.

III.1. China's Experience in Regulating AI Applications

The regulation of AI in China is going its own way. The efforts of lawmakers and legal scholars are aimed at finding ways to minimize the risks of AI applications, as well as finding a fair distribution of responsibility. In 2017, the State Council of the People's Republic of China published the Next Generation Artificial Intelligence Development Plan. The AI Development Plan is a national strategic-level law that establishes AI as a leading technology for the future. By 2030, it is planned that "the theory, technology and application of AI in China will generally reach the world level, making China a major global center of AI

innovation and achieving remarkable results in the development of the intelligent economy, which will be an important foundation for China to become a leading innovation and economic power."[2] At the same time, it was originally envisaged that by 2025, China would only "establish legal norms, ethical standards and an AI policy system, and form AI safety assessment and control capabilities." Widely known expert on Chinese law in Russia P.V. Troshchinsky (2015) always emphasizes that "the legislative policy of the Chinese authorities is characterized by the gradual development and subsequent adoption of local legislative acts necessary for the country." There is now a growing debate in China on the need to ensure the unity of legislation in this area, or even the preparation of a unified legal act on AI similar to the EU. The AI law has been included in the legislative work plan of the State Council of the People's Republic of China for 2023–2024. Regardless of the state's willingness to adopt a unified legal act, many lawyers believe that the legislation should establish a pluralistic regulatory system for AI, clarify the functions of national standards and industry standards, expand the space for group standards, and focus on the role of ethical standards (Song, 2024).

At the same time, Chinese lawmakers are very cautious in formulating final regulations and are moving in this direction in "small steps." Since 2022, China has started to consistently adopt regulations on algorithmic recommendations, deep synthesis technologies and generative AI services. The AI regulation was built based on its application areas. For example, the Regulation on the Management of Algorithmic Recommendations in Internet Information Services[3] (hereinafter referred to as the Algorithm Regulation) took effect on 1 March 2022 and introduced a regulation for the establishment of an algorithm registration system. Companies are required to apply for algorithm registration through the Algorithm Registration System of the Cyberspace Administration of China (CAC, 国家互联网信息办公室) and disclose the basic operating parameters of the technology.

[2] Available at: https://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm.

[3] 《互联网信息服务算法推荐管理规定》. Available at: https://www.gov.cn/gongbao/content/2022/content_5682428.htm [Accessed 10.12.2025].

The Regulation on the Management of Deep Fusion in Internet Information Services[4] of 25 November 2022 (hereinafter referred to as the Deep Fusion Regulation) was adopted to strengthen the comprehensive management of Internet information services, and it also provides a framework for the registration of AI systems. The registration procedures and rules are standardized based on practice. On 10 July 2023, the CAC issued the Interim Measures on the Management of Generative Artificial Intelligence Services[5] (hereinafter referred to as the Interim Measures) that took effect on 15 August 2023. This act defines the concept of generative AI and compliance obligations for providers of relevant products and services. In doing so, China has implemented a "dual registration" mechanism consisting of an algorithm registration system and an AI registration system (for large language models (LLM)). The registration of large language models requires the most effective interaction between regulators and service providers in order to gain regulatory experience and develop clear and specific rules that encourage companies to comply with their algorithm registration obligations, especially in terms of assessing the security of large language models. Disclosure of algorithms at the time of registration is helpful. According to regulators, the transparency of AI applications in specific areas should be audited.

In parallel with the formulation of national rules and regulations, local acts were introduced, such as the Regulations on the Development of Artificial Intelligence Industry in Shenzhen Special Economic Zone (《深圳经济特区人工智能产业促进条例》) dated 1 November 2022 or the Regulations on the Development of Artificial Intelligence Industry in Shanghai (《深圳经济特区人工智能产业促进条例》) dated 10 October 2022. That is, regional rulemaking was also included in addressing the applicability of specific AI products. The above-mentioned acts stipulate that the development of the AI industry follows the principles of technology advancement, application orientation, human-centered design, security and manageability. Initially, Chinese law did not

[4] 《互联网信息服务深度合成管理规定》. Available at: https://www.gov.cn/zhengce/zhengceku/2022-12/12/content_5731431.htmhttps://www.gov.cn/xinwen/2021-01/24/content_5582219.htm [Accessed 10.12.2025].

[5] 《生成式人工智能服务管理暂行办法》. Available at: https://www.gov.cn/zhengce/zhengceku/202307/content_6891752.htm [Accessed 10.12.2025].

treat AI as a separate object, but as part of the industry. In particular, the Shanghai Document mentioned above defined that the term "AI industry" refers to the key industries such as research, development and production of software and hardware products, system applications and integrated services related to artificial intelligence, as well as the integration and application of artificial intelligence in human life support, social management, economic development and other related fields (Art. 3 of the Regulation on the Development of the Artificial Intelligence Industry).

On 1 March 2024, China's National Cybersecurity Technology Standardization Committee officially released the Basic Security Requirements for Generative AI Services[6] (hereinafter referred to as the Basic Security Requirements). This law details the requirements for implementing the relevant provisions of the Interim Measures, such as the legitimacy of data sources, content security, etc., and provides effective methods for generative AI service providers to conduct security assessments in practice. It not only improves the internal security capabilities of companies in the field of generative AI services, but also provides standards for regulators to assess the security level of specific services.

The regulation of AI at the Statute level shall not conflict with the laws of the PRC that form the basis of the legal regulation of the country's information and technology development, such as PRC Cybersecurity Act,[7] PRC Data Security Act,[8] PRC Personal Data Protection Act,[9] PRC Science and Technology Progress Act,[10] as well as the Regulations on Security Assessment of Internet Information Services with Public

[6] 《生成式人工智能服务安全基本要求》. Available at: https://www.tc260.org.cn/front/postDetail.html?id=20240301164054 [Accessed 10.12.2025].

[7] 《中华人民共和国科学技术进步法》. Available at: https://www.gov.cn/xinwen/2021-12/25/content_5664471.htm [Accessed 10.12.2025].

[8] 《互联网信息服务算法推荐管理规定》. Available at: https://www.cac.gov.cn/202%202-01/04/c_1642894606364259.htm [Accessed 10.12.2025].

[9] 《中华人民共和国个人信息保护法》. Available at: https://www.cac.gov.cn/2021-08/20/c_1631050028355286.htm?eqid=b7a9c7a1000acbc7000000026465ed77 [Accessed 10.12.2025].

[10] 《中华人民共和国科学技术进步法》. Available at: https://www.gov.cn/xinw en/2021-12/25/content_5664471.htm.

Opinion Attributes or Social Mobilization Capabilities[11] (hereinafter referred to as the Security Assessment Regulations), the Regulations on Management of Information Services of Public Internet User Accounts.[12] All these regulations together form a closely related framework for the management of algorithms in various sectors, especially in the provision of generative AI services. In general, the AI regulation in Chinese Law is concerned with establishing several types of obligations: obligations related to supervision mechanisms; obligations related to algorithm training; obligations related to content management; and obligations of service providers to users.

III.2. Key Features of China's Approach to AI Regulation

Against the backdrop of the rapid expansion of artificial intelligence applications, China has been actively developing mechanisms to ensure the legal enforceability of the principle of transparency. In many respects, the principle of transparency is referred to because of its key importance for solving the problem of "black box" of algorithms (Zhang, 2024). Researchers note that the concept of transparency has two components: formal and substantive (Zhang, 2024). Formal transparency implies the disclosure of basic information about artificial intelligence, which makes the deployment and use of artificial intelligence unclassified. Substantive transparency is closely related to the AI interpretability, which emphasizes a meaningful explanation of the disclosed information in an understandable form, thereby breaking down knowledge barriers and making the relevant information truly recognizable. At first glance, this approach appears to be consistent with the European model. Article 4 of the Algorithm Regulation stipulates that the provision of algorithmic recommendation services shall follow the principles of fairness and honesty, openness and transparency. Service providers shall formulate and make public the rules of recommendation (Art. 7), optimize the transparency and interpretability of rules for searching, sorting, selecting, promoting and

[11] 《具有舆论属性或社会动员能力的互联网信息服务安全评估规定》. Available at: https://www.cac.gov.cn/2018-11/15/c_1123716072.htm [Accessed 10.12.2025].

[12] 《互联网用户公众账号信息服务管理规定》. Available at: https://www.gov.cn/xinwen/2021-01/24/content_5582219.htm [Accessed 10.12.2025].

displaying content (Art. 12), inform users about the principles, purpose and intentions of recommendation services as well as about the basic mechanisms of algorithmic recommendation services (Kharitonova and Tianfang, 2022).

At the same time, Chinese legal acts have their own specificities. The peculiarities of the Chinese approach are as follows. First, much AI research in China is conducted through the prism of the rule of law doctrine (Zhang, 2024). In this context, it seems logical to say in general that the principles of transparency and explainability should be carefully considered when implementing the rule of law: based on the consideration of existing technical conditions, different disclosure and explanation obligations should be imposed according to the different capabilities of regulators and the public. A second feature of the Chinese approach can be described as AI systems categorization. The scholarship suggests that there has been a paradigm shift in digital law in China from the initial individual risk prevention by lawyers to systematic risk management based on hierarchical and classified management, managing the entire life cycle of AI systems (Zhao and Zhou, 2024). This perspective can also be compared with the European one. As in the EU, algorithm classification and safety management are seen as the main tool for algorithm safety management in the future. However, even here we find distinctive features of the Chinese approach, based on the previous experience of rulemaking.

The Algorithms Regulation establishes general criteria for the classification of recommendation algorithm service providers based on the attributes of the service's public opinion or social mobilization capacity, the content category, the number of users, the degree of importance of the data processed by the algorithm recommendation technology, the degree of interference with user behavior, etc. The regulation of algorithms at the level of platform supervision provides for a hierarchical and classified management system and, for general-purpose platforms, the obligation to keep web logs and to cooperate with the State. For platforms that are opinion leaders or have social mobilization capabilities, a registration system will be introduced, and a proactive security assessment will be required. Entities providing online news and information services must obtain a license to provide

news and information services in accordance with the law. Regarding the application of the Algorithm Regulation, the benchmark standard for the classification of algorithms also depends on the degree of "sensitivity" of the data but will be changed to an indicator of the degree of "importance" of the data, which is most consistent with the standards of the Data Security Act. For example, the Interim Measures on the Management of Generative AI Services require that the provision and use of generative AI services comply with fundamental socialist values, implement effective measures to prevent discrimination based on ethnicity, belief, or country of origin, avoid the concentration of algorithmic power, and refrain from violating individuals' rights or the rights and interests of others in relation to personal information. These measures further mandate the enhancement of transparency in the operation and deployment of generative AI services.

In cases of high-risk illegal and offensive content relating to national security and public safety, the obligation for service providers to take preventive measures should be more stringent. In cases of illegal and offensive content relating only to the infringement of private rights characterized by a low risk the obligation for service providers to take preventive measures should be relatively less stringent (Yao and Li, 2023). Thus, for different levels of AI risk, China adopts implementation standards based on classification and scenarios and promotes the principle of transparency in two aspects: protecting users' and the public's right to information and ensuring the implementation of national regulatory powers. However, unlike the European approach, China takes the position that, given the existing technical conditions, it is necessary to differentiate the content requirements of the information to be disclosed and explained according to the different capabilities of regulators and the public, and to adopt graded and classified sub-scenarios of implementation standards for different risk levels of different AI.

A third important feature of the Chinese approach can be described as a commitment to require the application of AI within the framework of socialist morality and ethics. It is generally accepted that discrimination should not create algorithmic power and should not violate individual rights and the right to personal information of others. At the same time, a feature of China's approach is that the Interim

Measures on the Management of Generative AI Services requires that the provision and use of generative AI services should be in line with core socialist values and that effective measures should be taken to prevent the creation of ethnic, religious, national, etc. discrimination. The Algorithms Regulation sets standards for information services in Art. 6–10 that focus mainly on the control of illegal and unwanted information, in Art. 8 clearly states that "no algorithmic model shall be designed to cause users to become dependent on or excessively consume information in violation of laws and regulations or in violation of ethics and morality." The Algorithms Regulation also includes a prohibition to use algorithms to manipulate information and influence public opinion on the Internet. Thus, in addition to written laws and regulations, The Algorithms Regulation requires adherence to public morality and ethics, business and professional ethics, fairness and honesty, openness and transparency, scientific validity, and honesty and trustworthiness.

It seems that the violation of socialist core values and discriminatory content are classified as high-risk and require special attention and management. It is believed that the credible development of AI in China can be ensured by establishing and improving the examination system, supervision system and legal liability system.

III.3. A Brief Comparison in the Context of Russian Initiatives

A brief comparison can also be made with the Russian approach to AI regulation. While China's model is characterized by proactive ethical guidelines and detailed regulatory efforts to align AI development with socialist values, Russia's regulatory landscape remains more fragmented and strategic in nature. The key act in Russia today is the National Strategy for the Development of Artificial Intelligence for the Period up to 2030[13] (hereinafter Strategy 2030), which outlines the main

[13] Decree of the President of the Russian Federation dated 10 October 2019 No. 490 "On the development of artificial intelligence in the Russian Federation" (together with the "National Strategy for the development of artificial intelligence for the period up to 2030"). Sobranie zakonodatelstva Rissijskoj Federatsii [Collection of Legislation of the Russian Federation]. 2019. No. 41. Art. 5700.

principles of the development and use of AI technologies. The principles of regulation of technology development identified in the Russian legislation have not been deeply interpreted either in law enforcement practice or in the explanations of authorized bodies and are largely of a recommendatory nature. At the same time, the principles listed in Strategy 2030 partially overlap with several similar postulates reflected in ethical codes and recommendations in Russia and overseas.

Summarizing the experience of legal regulation of AI in China in comparison with the approaches of the European Union allows identifying some ideas that can be adopted in Russian legislative practice. First, both Chinese and European experience has shown that the unified AI legislation tends to the concept of comprehensiveness against the background of prudence and gradation of risk classification. At the same time, the unique experience of segmented regulation in China and, to a certain extent, the experimental-legal approach in Russia allow acting as cautiously as possible when creating a regulatory sandbox system, creating a rather narrowly differentiated regulation of AI systems. In any case, a system of control and monitoring of compliance with the requirements of the law on the use of AI will be introduced.

However, the jurisdiction and the bodies or commissions empowered to exercise such control operate according to different principles. At the same time, the Chinese approach to compliance with socialist morality and ethics in the application of AI, which is stricter than in the EU in the spirit of "developing science and technology for good" ("科技向善"), is in line with the tendencies of Russian law to protect traditional spiritual and moral values, culture and historical memory. The Chinese approach, on the other hand, tends to be rigid regarding the issue of censorship, which should not become the main trend of Russian law. The European approach seems closer to the Russian philosophy.

Some scholars have pointed out that the EU's approach of restricting the use of high-risk AI systems may lead to over-regulation and negatively affect the development of industries such as generative AI or biometrics. China's risk regulation scheme is essentially a shift from classifying risks for different products to classifying different risks arising from the same product (Arrieta et al., 2019), which is more appropriate for technologies such as generative AI. Thus, when

forming the Russian position on the content of legal regulation of AI applications, it is necessary to consider the positions of different countries to minimize risks and develop appropriate legal solutions in this area.

IV. Conclusion

The paper aimed to demonstrate that transparency is a pivotal yet highly complex principle in AI regulation, with profound implications for legal accountability, trust, and governance. As for the current major legislative landscapes, we can observe two highly diverging approaches. The European Union proposes individual safeguards, human oversight, and systemic accountability through a binding regulatory instrument, while China's framework embeds transparency within a broader ideological commitment to maintaining social harmony and state-sanctioned moral values, combining algorithmic explainability with a strong emphasis on governance and state oversight.

Though both initiatives are commendable as they are aiming to highlight ever-growing importance of transparency, there are still crucial legislative issues to assess. For instance, the divergence between the EU's and China's approaches signals a significant challenge for the future, namely, the absence of a universally shared understanding of transparency. This issue presents challenges not only with respect to ensuring compliance and guaranteeing access to effective legal remedies for affected parties, but also, at a more systemic level, poses a significant obstacle to the pursuit of global harmonization in AI governance. Achieving such harmonization is of critical importance: without reconciling societal, cultural, and regulatory divergences, the development of AI standards risks becoming fragmented, thereby undermining the coherence and effectiveness of cross-border cooperation, regulatory alignment, and oversight mechanisms in the AI domain. Demanding full uniformity would be highly naïve — even seemingly evident legal issues such as the protection of fundamental rights differ state by state or region by region (Escobar, 2024). Therefore, in the case of AI regulation, a regulatory field that can be described by extreme differences in regional developments and strategies, we can

hardly imagine a consolidated transparency regulation. Nonetheless, we propose and advocate that cooperative efforts should be made to commence a discussion on shared legal values and goals between authorities. In this regard, however, we also underline that transparency "alone," even when mandated, is insufficient. The inherent opacity of complex AI systems, the technical difficulties of achieving meaningful explainability and the varying capacities for regulatory enforcement necessitate a multifaceted approach. Future regulatory efforts must integrate technological innovation, dynamic monitoring mechanisms, and adaptive legal frameworks capable of responding to the evolving nature of AI. Lastly, the principle of transparency must be continually reimagined — not only as a legal obligation but also as an evolving societal expectation. Achieving this will require sustained interdisciplinary collaboration, international dialogue, and the political will to bridge competing regulatory philosophies, ensuring that transparency in AI is not merely a rhetorical ideal but a functional, enforceable reality.

References

Arrieta, A.B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R. and Herrera, F., (2019). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *In Information Fusion*, 58, pp. 82–115, doi: 10.1016/j.inffus.2019.12.012.

Birrer, A. and Just, N., (2024). What we know and don't know about deepfakes: An investigation into the state of the research and regulatory landscape. *New Media & Society,* doi: 10.1177/14614448241253138.

Busuioc, M., Curtin, D. and Almada, M., (2022). Reclaiming transparency: contesting the logics of secrecy within the AI Act. *European Law Open*, *2*(1), pp. 79–105, doi: 10.1017/elo.2022.47.

Castelvecchi, D., (2016). Can we open the black box of AI? *Nature*, 538(7623), pp. 20–23, doi: 10.1038/538020a.

Chaudhary, G., (2024). Unveiling the Black Box: Bringing Algorithmic Transparency to AI. *Masaryk University Journal of Law and Technology*, 18(1), pp. 93–122, doi: 10.5817/mujlt2024-1-4.

Diakopoulos, N. and Koliska, M., (2017). Algorithmic Transparency in the News Media. *Digital Journalism*, 5(7), pp. 809–828.

Durán, J.M. and Jongsma, K.R., (2021). Who is afraid of black box algorithms? On the epistemological and ethical basis of trust in medical AI. *Journal of Medical Ethics*, medethics-106820, doi. 10.1136/medethics-2020-106820.

Escobar, G.G., (2024). Fragmentation in the European and Inter-American human rights courts regarding the scope of religious autonomy: An analysis of the use of sources and methodologies. *Oxford Journal of Law and Religion*, 13(1), pp. 67–84, doi: 10.1093/ojlr/rwae022.

Golpayegani, D., Pandit, H.J. and Lewis, D., (2023). To Be High-Risk, or Not To Be — Semantic Specifications and Implications of the AI Act's High-Risk AI Applications and Harmonised Standards. *ACM International Conference Proceedings*, 12 June 2023, pp. 905–915, doi: 10.1145/3593013.3594050.

Gosztonyi, G. and Lendvai, G.F., (2023). Deepfake: A Multifaceted Dilemma in Ethics and Law. *Journal of Information Ethics*, 32(2), pp. 109–112, doi: 10.2307/JIE.32.2.109.

Gryz, J. and Rojszczak, M., (2021). Black box algorithms and the rights of individuals: no easy solution to the "explainability" problem. *Internet Policy Review*, 10(2), doi: 10.14763/2021.2.1564.

Helberger, N. and Diakopoulos, N., (2022). The European AI Act and How It Matters for Research into AI in Media and Journalism. *Digital Journalism*, 11(9), pp. 1751–1760, doi: 10.1080/21670811.2022.2082505.

Hosain, M.T., Anik, M.H., Rafi, S., Tabassum, R., Insia, K. and Siddiky, M.M., (2023). Path to Gain Functional Transparency in Artificial Intelligence with Meaningful Explainability. *Journal of Metaverse*, 3(2), pp. 166–180, doi: 10.57019/jmv.1306685.

Kharitonova, Y.S. and Tianfang, Y., (2022). Recommender Systems of Digital Platforms in China: Legal Approaches and Practices for Ensuring Algorithm Transparency. *Zakon*, 9, pp. 40–49, doi: 10.37239/0869-4400-2022-19-9-40-49.

Larsson, S. and Heintz, F., (2020). Transparency in artificial intelligence. *Internet Policy Review*, 9(2), pp. 1–16, doi: 10.14763/2020.2.1469.

Lendvai, G.F. and Gosztonyi, G., (2024). Deepfake y desinformación — ¿Qué puede hacer el derecho frente a las noticias falsas

creadas por deepfake? *IDP Revista De Internet Derecho Y Política*, 41, doi: 10.7238/idp.v0i41.427515.

Litwin, A.S. and Racabi, G., (2024). Varieties of AI Regulations: the United States perspective. *Industrial and Labor Relations Review*, 77(5), pp. 799–812, doi: 10.1177/00197939241278956a.

Mania, K., (2022). Legal Protection of Revenge and Deepfake Porn Victims in the European Union: Findings from a Comparative Legal Study. *Trauma Violence & Abuse*, 25(1), pp. 117–129, doi: 10.1177/15248380221143772.

Maslova, E.A. and Sorokova, E.D., (2022). The Dialectics of Ethics and Law in the Regulation of Artificial Intelligence: Case of the EU. *Sovremennaya Evropa*, 5, pp. 19–33, doi: 10.31857/S0201708322050023. (In Russ.).

Miller, T., (2018). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267, pp. 1–38, doi: 10.1016/j.artint.2018.07.007.

Moreno, F.R., (2024). Generative AI and deepfakes: a human rights approach to tackling harmful content. *International Review of Law Computers & Technology*, 38(3), pp. 297–326, doi: 10.1080/13600869.2024.2324540.

Mourby, M., Cathaoir, K.Ó. and Collin, C.B., (2021). Transparency of machine-learning in healthcare: The GDPR & European health law. *Computer Law & Security Review*, 43, p. 105611, doi: 10.1016/j.clsr.2021.105611.

Novelli, C., Casolari, F., Rotolo, A., Taddeo, M. and Floridi, L., (2023). Taking AI risks seriously: a new assessment model for the AI Act. *AI & Society,* doi: 10.1007/s00146-023-01723-z.

Novelli, C., Gaur, A. and Floridi, L., (2025). Two Futures of AI Regulation under the Trump Administration. *SSRN preprint*, pp. 1–13. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5198926 [Accessed 10.12.2025].

Papadouli, J.S., (2022). Transparency in Artificial Intelligence: A Legal Perspective. *Journal of Ethics and Legal Technologies*, 4(1), pp. 24–40.

Pošćić, A. and Martinović, A., (2023). Regulatory Sandboxes under the Draft EU Artificial Intelligence Act: An Opportunity for SMEs.

InterEULawEast Journal for the International and European Law Economics and Market Integrations, 9(2), pp. 71–117, doi: 10.22598/iele.2022.9.2.3.

Pu, K. and Xiang, W., (2023). Opportunities and challenges in using ChatGPT as generative AI and response strategy. *Journal of Chongqing University*, 29(3), pp. 102–114.

Raja, A.K. and Zhou, J., (2023). AI Accountability: Approaches, Affecting Factors, and Challenges. *Computer*, 56(4), pp. 61–70, doi. 10.1109/mc.2023.3238390.

Shen, F., (2024). Risk and management of generative artificial intelligence: solving the "Collingridge's dilemma." *Journal of Zhejiang University*, 6, pp. 73–91.

Söderlund, K., Engström, E., Haresamudram, K., Larsson, S. and Strimling, P., (2024). Regulating high-reach AI: On transparency directions in the Digital Services Act. *Internet Policy Review*, 13(1), doi: 10.14763/2024.1.1746.

Song, H., (2024). Development of normative structure in artificial intelligence legislation. *Journal of East China University of Political Science and Law*, 5.

Tartaro, N., (2023). Regulating by standards: current progress and main challenges in the standardisation of Artificial Intelligence in support of the AI Act. *European Journal of Privacy Law & Technologies*, 1, pp. 147–174, doi: 10.57230/ejplt222at.

Troshchinsky, P.V., (2015). Development of Chinese legislation in recent years. *International Public and Private Law,* 3, pp. 36–40.

Yao, Z.W. and Li, Z.L., (2023). Legal risk regulation of generative artificial intelligence content. *Journal of Xi'an Jiaotong University*, 43(5), pp. 147–160. (In Russ.).

Vesnic-Alujevic, L., Nascimento, S. and Pólvora, A., (2020). Societal and ethical impacts of artificial intelligence: Critical notes on European policy frameworks. *Telecommunications Policy*, 44(6), p. 101961, doi: 10.1016/j.telpol.2020.101961.

Walmsley, J., (2020). Artificial intelligence and the value of transparency. *AI & Society*, 36(2), pp. 585–595, doi: 10.1007/s00146-020-01066-z.

Zhao, C. and Zhou, J., (2024). On the paradigm shift of digital legal research: systemic risk management. *Qiushi Academic Journal*, 4(136).

Zhang, Y., (2024). On the legal realization of the principle of transparency in the field of artificial intelligence. *Policy and Legal Discussion Series*, 2. p. 124.

Information about the Authors

Yulia Kharitonova, Research and Education Center for Legal Studies of Artificial Intelligence and Digital Economy, Lomonosov Moscow State University, Moscow, Russian Federation
sovet2009@rambler.ru (Corresponding Author)
ORCID: 0000-0001-7622-6215

Gergely Ferenc Lendvai, Department of Public Administration & Digital Authoritarianism Research Lab, ELTE, Ludovika University of Public Service, Budapest, Hungary
ORCID 0000-0003-3298-8087

Received 26.11.2024
Revised 12.12.2024
Accepted 07.01.2025