International Education Forum, 2025, Volume 3, Issue 8
http://ojs.bbwpublisher.com/index.php/IEF
Online ISSN: 2981-8605
Print ISSN 3083-4902

Research on the Application of the Transparency Principle in the EU Artificial Intelligence Act

Mengting Zhu*

School of Law, Guizhou Normal University, Guiyang 550025, Guizhou, China

**Author to whom correspondence should be addressed.*

Abstract: The EU Artificial Intelligence Act establishes the principle of transparency to ensure the traceability, interpretability, and communicability of artificial intelligence (AI) systems. This principle is reflected to varying degrees in AI systems of different risk levels, with particularly stringent requirements for high-risk AI systems. However, in practical application, the transparency principle faces multiple challenges, such as conflicts with trade secret protection, technical complexity, and ambiguous liability definition. Although the EU has adopted a series of specific provisions to ensure the implementation of the transparency principle, including requirements for information disclosure and documentation, these measures still need to further balance the interests of all parties during implementation to achieve the goal of promoting the healthy development of AI technology while protecting public rights and interests, social fairness, and justice.

Keywords: EU Artificial Intelligence Act; Transparency principle; Application issues

Online publication: September 18, 2025

1. Introduction

In recent years, with the vigorous development of artificial intelligence, its application in various fields has not only profoundly transformed people's work and life but also raised issues such as algorithmic discrimination, privacy infringement, and ambiguous liability attribution, which have made the public lack necessary confidence in AI. To address this, the European Commission officially published the Artificial Intelligence Act in March 2021, aiming to build a consistent regulatory framework compatible with EU common values [1]. As an important basic principle of the EU Artificial Intelligence Act, the transparency principle can effectively overcome technical difficulties. By focusing on system traceability, interpretability, and communicability, it promotes public understanding and trust in technology and enhances the development level of technology. However, in practice, the transparency principle often faces many difficulties, and how to balance the legitimate interests of enterprises, technical barriers, and other issues has become one of the focuses in the field of AI at this stage.

2. Specific application of the transparency principle in the EU Artificial Intelligence Act

2.1. Application in AI systems of different risk levels

According to different risk levels, the EU Artificial Intelligence Act classifies AI systems into four categories: unacceptable risk, high risk, limited risk, and minimal risk. The transparency principle applies to systems of different risk levels to varying degrees. AI systems with unacceptable risks are explicitly prohibited from entering the market, and the transparency principle does not apply to them. High-risk systems are subject to the "absolutely acceptable transparency" principle, and strict and comprehensive transparency requirements must be reflected in the qualification requirements for high-risk systems to enter the market. AI systems must meet the technical documentation requirements specified in Article 11 of the Artificial Intelligence Act and satisfy all requirements set forth in Articles 8-15 of the Act [2]; the system must be recognized by relevant parties capable of reasonably interpreting and applying output results, and it is stipulated that the ability must be reasonably designed to facilitate the performance of duties by providers and users; it must also be accompanied by a complete user manual including accurate and comprehensive information about features, functions, and all possible usage boundaries; the system must have the ability to automatically log all important events, and the event records must be detailed enough to facilitate supervision and verification. AI systems with limited risks shall fulfill their obligations of information disclosure and transparency. Even if the risk is very low, relevant affected parties must be informed transparently [3].

2.2. Requirements for AI system providers and operators

In the EU Artificial Intelligence Act, both providers and operators have numerous transparency-related obligations. For providers, general-purpose AI model providers shall formulate and maintain technical documentation and provide information and documents to ensure transparency. For general-purpose AI models with systemic risks, providers shall conduct model evaluation and carry out adversarial testing, which shall be recorded in documents to identify and eliminate systemic risks. For example, providers of general-purpose AI models shall comply with transparency guidelines, take appropriate measures to ensure that no content violating EU laws is generated, and disclose data involving copyright used in training models. Operators also have numerous transparency obligations to fulfill. When operators interact directly with natural persons, they must ensure interaction with the AI system; when operating systems that generate synthetic information, operators shall require the output to be in a machine-readable format to ensure that synthetic content can be detected and identified as machine-generated or manipulated; when operating emotion recognition systems or biometric classification systems, operators shall inform contacts and process personal data; when operating systems that generate or manipulate deepfake content, operators shall disclose that the deepfake content is generated or manipulated by AI; when using systems to generate or modify text information involving public interest issues, disclosure shall be made, etc. [4].

3. Problems in the application of the transparency principle

3.1. Conflict with trade secret protection

For AIGC companies, algorithms, data processing methods, and training data are important outputs of R&D investment and constitute key elements of trade secrets. Taking the unique algorithms and training data of image processing companies as an example, they are core to the company's market competitive advantage.

However, in accordance with the transparency principle, such companies need to disclose summaries of training data, etc., which increases the risk of trade secret leakage and may weaken their competitiveness. For start-ups with insufficient resources, trade secrets are the foundation of their survival, and being forced to disclose core technologies may bring a survival crisis; large high-tech companies are also reluctant to disclose trade secrets, as this may weaken their core competitiveness [5].

3.2. Implementation difficulties caused by technical complexity

Due to the high complexity of artificial algorithms, there are significant difficulties in the practical application of AI technology. Taking deep learning algorithms as an example, the operation model of such AI algorithms is trained with a large amount of sample data, and finally optimized and updated during the training process, and decision-making is based on training data. The operation process is cumbersome and complex, making it difficult to provide a simple explanatory analysis. For example, the AI algorithm model for deep learning image processing has numerous layers of operations, neuron nodes, and weighted changes. Changes in each neuron node will cause significant changes in results, making it almost impossible to provide explanatory analysis to ordinary users. The responsiveness of legal norms related to algorithm transparency does not match the highly complex characteristics of algorithms, making it difficult to effectively set clear and feasible standards for algorithm transparency systems.

3.3. Ambiguity in liability definition

The transparency of AI system decisions caused by algorithm operation among algorithm design, data provision, training models, human-computer interaction, and other steps makes it more difficult to allocate responsibilities. For example, it is difficult to distinguish whether the liability for a self-driving car accident is caused by data issues, algorithm problems, or human factors, which may easily lead to mutual shirking of responsibilities. The new technological revolution continues, resulting in delayed legislation, and there is still a lack of global standards worldwide. Due to different technical standards and regulations among countries, cross-border application becomes more difficult. The disconnection between technology and standards increases the difficulty of implementing the transparency principle, making it inconvenient to accurately identify responsible subjects and difficult to fully protect the interests of victims and guide technological development.

4. Suggestions for solving the application problems of the transparency principle

4.1. Measures to balance transparency and trade secret protection

At the legislative level, the Artificial Intelligence Act and supporting regulations should be refined to clarify the “necessary limits” of transparency disclosure. Drawing on the classified and hierarchical disclosure mechanism, core information can be divided into “mandatory disclosure items”, “limited disclosure items”, and “confidential items.” For example, the decision logic framework and data source categories of high-risk systems must be publicly disclosed, while core trade secrets such as specific algorithm parameters can be kept confidential but must be encrypted and filed with regulatory authorities. At the same time, exceptions to trade secret protection should be clarified. When decisions may cause significant harm, regulatory authorities have the right to require temporary disclosure and sign confidentiality agreements.

An independent committee composed of technical experts, legal experts, and industry representatives should be established to conduct specialized reviews of enterprises’ applications for trade secret protection.

Applicants for protection shall submit applications stating the reasons for protection and alternative disclosure methods, such as simple technical explanations or third-party inspections. The committee shall decide whether to grant protection within 15 working days and specify alternative disclosure methods or conduct regular inspections of confidential information [6].

Enterprises should establish a full-process information classification management system and mark public, internal shared, and core confidential information during the development stage. For information to be disclosed, "desensitized disclosure" should be adopted, such as disclosing the source field of training data rather than samples, and showing the weight trend of decision-influencing factors rather than specific parameters. A special information disclosure management department should be established, with technical, legal, and compliance personnel jointly reviewing the content.

Industry associations should be urged to issue guidelines on Transparency and Trade Secret Protection to determine disclosure norms and confidentiality limits [7]. For example, text-to-image AI should disclose data copyright compliance certification reports and technical principle summaries, while keeping key technologies confidential. Trade secret protection certification and industry-shared technology platforms should be established to provide services such as encrypted packaging and reduce the compliance costs for small and medium-sized enterprises.

4.2. Strategies to address technical complexity

Addressing technical complexity requires joint efforts of technological innovation and institutional innovation. Increase investment in explainable AI during technology R&D and develop corresponding explainable tools for different application scenarios. For "black-box" deep learning, develop model interpretability technologies to convert complex deep learning calculations into intuitive causal relationship diagrams. For example, AI products in the medical field use visualization methods to display key diagnostic data and weights. Develop modular algorithms, restrict high-risk systems to adopt detachable designs, and the functions and logic of each part of the algorithm are explainable. In the EU, special funds can be established to encourage research in this area, and rewards can be provided for technologies that achieve breakthroughs [8].

Develop transparency operation guidelines. From the perspective of the market and supervision, form flexible implementation guidelines for the transparency principle. According to the content of transparency implementation guidelines issued for different fields, there is generally room for annual revisions to incorporate newly emerging technologies. For example, for facial recognition technology, it is recommended to disclose key technical performance parameters and accuracy; for unmanned driving technology, it is recommended to record and retrieve decision parameters. Expressions should be "result-oriented" to avoid detailed provisions on technical details, and expert groups should review and revise them every 2 years [9].

Build a multi-participation technical verification mechanism and introduce third-party institutions to evaluate and certify the degree of transparency implementation. Verification institutions verify the authenticity of information disclosed by enterprises through technical testing, such as testing whether algorithm fairness is consistent with disclosed goals, and make the results public. The EU should establish an institution qualification certification system, clarify access norms, and penalty measures for violations.

Strengthen technical education and popular science, develop explanatory tools and materials for different groups, such as the "AI Decision Explanation Manual" for the public and technical training for supervisors. Incorporate basic AI courses into primary and secondary schools to cultivate public literacy. Encourage

enterprises to develop “user-friendly” interfaces and provide basic or professional hierarchical explanation functions.

4.3. Methods to clarify the liability definition

Building a liability identification mechanism requires innovations in the rule of law, mechanisms, and technology. When revising the Artificial Intelligence Act, a “full-life-cycle liability identification” model should be established to identify responsible subjects from three stages: design, R&D, deployment, and operation. Algorithm developers are responsible for algorithm safety and interpretability; data providers are responsible for data authenticity and compliance; deployers are responsible for scenario matching; operators are responsible for daily operation supervision. Implement the “liability ladder” principle, where liability weights are graded based on technical control capabilities and profit levels. In the fully autonomous driving mode, providers bear primary liability. If they fail to perform their duties without special reasons, they shall bear a certain proportion of liability, and circumstances such as unforeseeable technical vulnerabilities that mitigate legal liability should be specified.

Establish a liability identification standard committee involving multiple parties to formulate detailed standards and guidelines, and formulate judgment rules for high-risk application scenarios of medical AI. For example, AI diagnostic errors should consider the scope of use and the review of medical personnel. Extract key elements from typical cases and update them regularly to adapt to technological changes.

Implement third-party liability certification and traceability, promote blockchain full-life-cycle traceability, and record design parameters, data sources, and other information for accident traceability. For example, autonomous driving records software updates and decision data; develop intelligent liability identification to simulate liability proportions; the EU mandates that high-risk systems be connected to a unified platform and formulate technical and safety specifications.

Build a diversified dispute resolution mechanism, establish AI dispute mediation centers, special arbitration institutions, and specialized court tribunals, promote liability insurance systems, and insurance institutions set premium rates based on transparency and risks to encourage enterprises to improve their standards [10].

5. Conclusion

The transparency principle in the EU Artificial Intelligence Act plays a pivotal role in promoting technological development in the field of AI and protecting public interests. It has targeted and detailed application provisions for AI systems with different risk levels and provides detailed transparency requirements for providers and operators. However, in practical application, the transparency principle still faces problems such as balancing with trade secret protection, difficulties in implementation due to technical complexity, and an ambiguous liability definition. Reasonable measures can overcome these problems to a certain extent. However, it is worth noting that with the progress and development of AI technology, other problems will inevitably arise. For example, the transparency principle should be continuously considered and studied, and relevant policies and regulations should be continuously revised and improved to ensure their sound and effective implementation in the field of AI, promote the development of AI technology in a transparent and trustworthy environment, and achieve positive interaction between technological progress and social fairness, justice, and protection of public interests.

Disclosure statement

The author declares no conflict of interest.

References

[1] Li XS, Li TK, 2025, On the Legal Risk Regulation of Artificial Intelligence Systems—Taking the EU Artificial Intelligence Act as an Example. Journal of University of Science and Technology Beijing (Social Sciences Edition), 41(1): 86–102.

[2] Han R, 2025, Copyright Challenges and Responses in the Era of Generative Artificial Intelligence—With Reference to EU Legislation. Journal of Zhaoqing University, 46(2): 45–53.

[3] He ZH, 2025, The Product Safety Approach to Artificial Intelligence Legislation—A Critical Interpretation of the EU Artificial Intelligence Act. SJTU Law Review, 2025(1): 153–164.

[4] Lu CY, 2025, EU Artificial Intelligence Strategy and Prospects for China-EU Cooperation in Artificial Intelligence Governance. Contemporary World, 2025(5): 31–36.

[5] Liao XJ, Shi LD, 2025, Analysis of National Security Risks of the EU Artificial Intelligence Act from the Perspective of Externality. Science and Technology Management Research, 45(1): 238–245.

[6] Du J, 2025, Innovation of EU Citizens' Digital Competence Framework in the Context of Artificial Intelligence and Its Enlightenment. Journal of Academic Library and Information Science, 43(2): 131–139.

[7] Gao ZH, Zhang XZ, 2025, Analysis of Policy Tools, Path Characteristics and Influences of the EU's Participation in Global Artificial Intelligence Governance. Forum on Science and Technology in China, 2025(2): 150–160.

[8] Zou J, Ji CY, 2025, The EU's Approach to AI Risk Regulation and Its Enlightenment—Based on the Interpretation of the Artificial Intelligence Act. Journalism Research, 2025(1): 31–44 + 118.

[9] Huang HY, Yang X, 2025, Interpretation of the EU Artificial Intelligence Regulatory System Based on the Artificial Intelligence Act. Library Development, 2025(1): 12–24.

[10] Li KL, 2025, The Response and Effectiveness of the Artificial Intelligence Act to the Construction of EU Digital Sovereignty. Journal of Political Science and Law, 2025(2): 156–169.