© IJARW | ISSN (O) - 2582-1008
May 2026 | Vol. 7 Issue. 11
www.ijarw.com

PROHIBITED ACTS IN ARTIFICIAL INTELLIGENCE ACTIVITIES AND THEIR SIGNIFICANCE FOR PREVENTING SOCIAL RISKS FROM A LEGAL PERSPECTIVE

Nguyễn Thị Thùy Giang

Department of Law, TNU- University of Sciences, Thai Nguyen 25000, Vietnam

ABSTRACT

This article analyzes prohibited acts in artificial intelligence activities under the 2025 Law on Artificial Intelligence, with a focus on Article 7 on prohibited acts. Using doctrinal analysis, comparative legal analysis, and reference to selected real-world developments, the article clarifies the grounds for establishing such prohibitions, their legal characteristics, their boundaries of application, and their significance for preventing social risks. The article argues that the core value of these provisions lies not only in post-violation punishment, but also in establishing legal limits on forms of AI use that may erode human autonomy, obscure accountability, undermine public trust, or generate systemic risks to social order. By comparing Vietnam's approach with the European Union's AI Act, the Council of Europe Framework Convention on AI, OECD recommendations, and recent guidance from European data protection authorities, the article proposes several recommendations to clarify the legal indicia of prohibited conduct, strengthen data governance, require impact assessments for high-risk AI systems, and improve mechanisms for transparency, traceability, and inter-agency enforcement coordination.

Keyword: artificial intelligence; prohibited acts; prevention of social risks; personal data; transparency; accountability.

1. INTRODUCTION

Artificial intelligence is evolving from a supportive technology into one capable of directly shaping decisions, allocating opportunities, influencing behavior, and generating content that affects social life. This makes the legal risks of AI different from those of traditional information technology: risks may arise from training data, system design, deployment methods, the degree of automation, and the capacity for large-scale dissemination. Accordingly, AI law cannot stop at addressing consequences after the fact; it must establish in advance certain "red lines" for particularly dangerous practices.

On 10 December 2025, the National Assembly passed the Law on Artificial Intelligence as a framework statute aimed at regulating AI outputs, patterns of use, and social risks, rather than intervening too deeply in the underlying technological model [1], [2]. Within that structure, Article 7 on prohibited acts plays a central role, because it is where principles such as human-centeredness, the protection of human rights, data protection, transparency, and accountability are translated into directly applicable behavioral limits.

At the international level, current AI governance trends combine innovation promotion with risk control. The European Union's AI Act prohibits AI practices that present an "unacceptable level of risk," such as harmful manipulation and the exploitation of vulnerabilities of vulnerable groups [3]. The Council of Europe Framework Convention on AI emphasizes that the entire lifecycle of AI systems must remain compatible with human rights, democracy, and the rule of law [4]. The OECD likewise treats AI risks not merely as technical risks but as social risks, including bias, privacy violations, information security failures, and information manipulation [5], [6]. Studying Article 7 therefore has significance not only for interpreting domestic law, but also for situating Vietnam's AI governance model within a comparative legal framework.

This article addresses three questions: (i) what legal grounds justify the establishment of prohibited acts in AI activities; (ii) how these categories of prohibited conduct should be

understood and applied; and (iii) what significance these prohibitions have for preventing social risks and for improving AI law in Vietnam. The article combines doctrinal analysis, comparative legal analysis, and case-study methodology through selected international examples involving deepfakes, voice cloning, training data, and "AI washing."

2. LEGAL OVERVIEW OF PROHIBITED ACTS IN ARTIFICIAL INTELLIGENCE ACTIVITIES

2.1. Grounds for Establishing Prohibited Acts in the Law on Artificial Intelligence

In legal theory, a prohibition is a regulatory technique used to identify conduct that legal subjects are not permitted to engage in because it threatens values protected by law. In the case of AI, the role of prohibitions is particularly prominent because this technology has three features that heighten risk: a high degree of automation, inferential capacity over large datasets, and the ability to amplify consequences on a broad scale.

The 2025 Law on Artificial Intelligence establishes prohibited acts on the basis of core governance principles such as human-centeredness, the protection of human rights, privacy, the public interest, national security, transparency, non-discrimination, meaningful human oversight, and accountability [1]. Article 7 is therefore not a random collection of prohibitory clauses, but a concretization of the foundational values the law seeks to protect.

The grounds for establishing prohibitions also stem from the lifecycle-based risk characteristics of AI. The legality of a system cannot be assessed solely by looking at its final output; it must also take into account the data source, the training process, the control mechanisms, the deployment method, and the capacity for remediation when incidents occur. This is also the general logic of modern AI governance frameworks: controlling risk early and throughout the system lifecycle, rather than intervening only once harm has materialized [3], [4], [6].

2.2. Legal Characteristics of Prohibited Acts in Artificial Intelligence Activities

Prohibited acts in AI activities have four main legal characteristics. First, they are anticipatory-risk acts: the law may intervene even before harm occurs when the risk of serious damage has already become apparent. Second, prohibited conduct spans the entire AI lifecycle, from data collection, training, and testing to deployment and use. Third, these acts are clearly interdisciplinary in nature, requiring simultaneous consideration of data law, cybersecurity law, civil law, consumer protection law, intellectual property law, and criminal law. Fourth, the protective focus of these prohibitions is directed strongly toward human rights, vulnerable groups, and the public interest. This reflects the logic of modern AI governance: technology is legitimate only when placed within the limits of human dignity, equality, freedom of information, and accountability.

3. LEGAL ANALYSIS OF PROHIBITED ACTS IN ARTIFICIAL INTELLIGENCE ACTIVITIES

3.1. Misappropriating or Seizing Artificial Intelligence Systems to Commit Unlawful Acts

Clause 1, Article 7 prohibits the exploitation or seizure of artificial intelligence systems for the purpose of committing unlawful acts. In essence, this category covers conduct that uses AI as a tool to amplify illegality, or that infiltrates and takes control of AI systems in order to serve unlawful purposes. In such cases, unlawfulness does not arise merely because technology is involved; rather, it arises because AI increases the scale, speed, camouflage, and spread of the underlying violation. This prohibition therefore affirms that the right to research, access, and use technology cannot be invoked to justify the commission of the underlying illegal act.

From an application perspective, this category includes conduct such as taking over administrator accounts to alter the outputs of an AI system; using generative AI to distribute phishing emails on a large scale; or exploiting voicebots and chatbots to impersonate public authorities, banks, or relatives for the purpose of obtaining property unlawfully. The central legal issue is that AI does not change the nature of the unlawful act, but it does significantly change the degree of social danger posed by that act. For this reason, placing such conduct in the category of prohibited acts under sector-specific legislation is justified, because it allows the State to intervene at an early stage in the formation of risk.

3.2. Impersonation, Manipulation of Cognition, and Serious Harm to Human Beings

One of the most significant categories of prohibited conduct is the development, provision, deployment, or use of AI systems to impersonate, manipulate human cognition, behavior, or decision-making processes in ways capable of causing serious harm. The core legal issue here is the erosion of individual autonomy. Modern law protects human beings not only from tangible material harm, but also from the loss of access to truthful information, from decision-making that is no longer based on free and informed cognition, and from psychological-technological techniques designed to neutralize their ability to protect themselves.

The important point is that the law does not prohibit every form of simulation or every persuasive technology. The legal boundary is crossed only where there is impersonation or deception; where the conduct is systematic and intentional; where it substantially impairs cognitive capacity or autonomy; and where it leads to, or creates a high risk of, serious harm. This approach is close to that of the EU AI Act, which prohibits manipulative or deceptive techniques likely to cause significant harm [3]. Practice has already shown that this risk is real: in February 2024, the U.S. FCC confirmed that robocalls using AI-generated voices constitute an "artificial voice," thereby treating voice cloning used in scams as unlawful; the FTC likewise emphasized the need to develop preventive and detection solutions for abuses involving voice cloning [11]. In the Vietnamese context, conduct such as faking a relative's voice to request a money transfer, creating deepfake images to defame a person's reputation, or using chatbots to lure users into high-risk transactions should all be examined through the lens of this category of prohibited conduct.

3.3. Exploiting the Vulnerabilities of Vulnerable Groups

Point c, Clause 2, Article 7 prohibits exploiting the vulnerabilities of vulnerable groups, such as children, older persons, persons with disabilities, or persons who have difficulties with cognition or self-control. From a legal perspective, this is not merely a humanitarian provision, but also a mechanism for addressing structural asymmetry between the party controlling technology and the party affected by it. In AI environments, developers or deployers can analyze behavioral data, infer psychological states, identify moments of susceptibility, and optimize persuasive strategies for each user. As a result, the merely formal "consent" of vulnerable persons does not always reflect actual cognitive capacity or real ability to protect themselves.

The EU AI Act also prohibits exploiting vulnerabilities related to age, disability, or socioeconomic circumstances in ways that materially distort behavior and are likely to lead to significant harm [3]. This comparative experience reinforces the approach taken by the Law on Artificial Intelligence. In practice, unlawful conduct may take the form of learning applications that repeatedly prompt children to make purchases through virtual characters; insurance-sales chatbots targeting older persons with messages that exploit fears of illness; or AI interaction systems that induce persons with cognitive difficulties to enter into financial transactions. The common feature of these cases is that AI is not merely personalizing services, but exploiting human vulnerability.

3.4. Creating or Disseminating Fabricated Content That Causes Serious Harm to National Security, Public Order, or Social Safety

The Law on Artificial Intelligence prohibits creating or disseminating fabricated content capable of causing serious harm to national security, public order, or social safety. The legal harm here lies not only in the falsity of the information, but also in the capacity of AI-generated content to fracture public trust, generate social panic, affect crowd behavior, or interfere with important social processes. Compared with conventional fake news, fabricated content generated by AI is markedly more dangerous because it can appear highly realistic, is cheap to produce, and can be distributed at scale.

Many legal systems have shifted from a mindset of simply "combating fake news" to one of "governing AI-manipulated content." In the EU, Article 50 of the AI Act imposes transparency obligations for deepfakes and content generated or manipulated by AI; the European Commission is developing a Code of Practice on marking and labeling AI content to support compliance [7], [8]. This suggests that, when applying Article 7 in Vietnam, one should not stop at asking whether content is fake, but should instead evaluate the purpose of use, the context, the affected groups,

the scope of dissemination, and the potential for dangerous cascading effects.

3.5. Collecting, Processing, or Using Data Unlawfully

Clause 3, Article 7 prohibits the collection, processing, or use of unlawful data to develop, train, test, or operate AI systems. This is a structurally important provision for the entire statute, because data are the legal raw material of AI. If input data are collected unlawfully, processed for the wrong purpose, exceed what is necessary, or violate the rights of data subjects, then the legality of the AI system is called into question from the outset, regardless of how useful its output may appear.

Opinion 28/2024 of the EDPB on data protection aspects related to AI models emphasizes the need to examine closely the conditions under which a model may be considered anonymous, the validity of the legal basis for data processing during the development stage, and the consequences of unlawful data processing for subsequent deployment [9]. CNIL's recommendations likewise indicate that training AI on personal data is not absolutely prohibited under the GDPR, but is lawful only where the purpose, legal basis, scope of necessary data, information obligations, and mechanisms for the exercise of individual rights are clearly defined [10]. From the perspective of application in Vietnam, this suggests that competent authorities should ask not only whether a system performs well, but also whether its input-data chain is lawful, minimized, and verifiable.

3.6. Obstructing, Disabling, or Distorting Mechanisms for Human Oversight and Control

Clause 4, Article 7 prohibits obstructing, disabling, or distorting mechanisms for human oversight, intervention, and control over AI systems. In essence, this provision protects the principle of meaningful human oversight over technologies capable of significantly affecting individual rights and interests. If a system operates as a closed system, without points of intervention, without the ability to stop it, or without clarity as to who bears ultimate responsibility, the entire mechanism for protecting rights is weakened.

The EU AI Act likewise emphasizes that high-risk systems must be designed in a way that allows humans to supervise them, understand warnings, intervene, or deactivate them when necessary [3]. In practice, violations may take the form of deliberately disabling warning functions; designing interfaces so that operators cannot understand or cannot stop automated decisions; or misrepresenting the system so that managers believe a control mechanism still exists when in fact the system is fully closed. Such conduct is especially dangerous because it conceals risk precisely at the point that should have served as the system's "safety valve."

3.7. Concealing Information That Must Be Transparent or Accountable, or Falsifying Warning Labels

Clause 5, Article 7 prohibits concealing information that must be publicly disclosed, transparent, or subject to explanation; and erasing or falsifying warning labels, identifiers, or recognition information for content generated by AI. Legally speaking, transparency does not mean disclosing all technological secrets, but rather disclosing enough for individuals to know when they are interacting with AI, to know which content has been generated or manipulated by AI, to understand the system's principal risks, and to identify who is responsible when disputes arise.

When mandatory information is concealed or warning labels are falsified, users can no longer make informed decisions, regulators face difficulties in conducting ex post oversight, and victims encounter major obstacles in proving harm. In the context of rapidly developing generative AI, labeling and provenance obligations become important legal tools. The European Commission is developing a Code of Practice on marking and labeling AI content under Article 50 of the AI Act, emphasizing measures such as watermarking, metadata, and machine-readable markers [7], [8]. In addition, the practice known as "AI washing" shows that transparency in AI is not merely an ethical obligation: in 2024, the SEC sanctioned two investment advisory firms for making misleading statements about their use of AI [12].

3.8. Misusing Research, Testing, Evaluation, or Certification Activities to Commit Unlawful Acts

Clause 6, Article 7 prohibits misusing research, testing, evaluation, or certification activities involving AI systems to commit unlawful acts. This provision matters because the law must always balance freedom of research with the need to

protect society. Sandboxes, controlled testing, safety evaluation, and technology certification are necessary for innovation, but these "flexible" spaces can also be abused as a cover for unlawful data collection, dissemination of harmful content, or experimentation on users without a clear legal basis.

The key point is that research and testing do not negate the unlawfulness of conduct if the conduct in substance still infringes the rights and lawful interests of individuals or the public interest. An AI experiment conducted on personal data without a legal basis, a certification process used to spread malware, or a research project using deepfakes to manipulate public perception cannot be justified merely by invoking a "research purpose." This is also the general spirit of international AI governance frameworks [3], [4], [6].

4. THE SIGNIFICANCE OF PROHIBITIONS IN THE LAW ON ARTIFICIAL INTELLIGENCE FOR PREVENTING SOCIAL RISKS

4.1. Protecting Human Rights and Civil Rights in the Digital Environment

The first and most important significance of the prohibitions is the protection of human rights against new forms of harm created by AI. In the digital environment, the rights to privacy, reputation, equality, truthful information, and freedom from manipulation are threatened not only by direct human conduct, but also by systems capable of prediction, simulation, automation, and mass dissemination. By prohibiting practices such as impersonation, manipulation of cognition, exploitation of vulnerable groups, unlawful data use, and the disabling of human control, the Law on Artificial Intelligence creates an important layer of legal protection for individual dignity and autonomy.

4.2. Preventing Systemic Risks to Public Order, Social Safety, and National Security

The prohibitions also have special significance in preventing risks at the systemic level. The OECD regards AI incidents involving bias, privacy intrusions, misinformation, safety failures, and security failures as present realities rather than hypothetical risks [5]. With generative AI, a single violation can quickly turn into a large-scale social risk because reproduction costs are low and dissemination can occur extremely rapidly. Prohibiting particularly dangerous AI practices is therefore a way of establishing a "safety valve" for both the legal system and the social information environment.

4.3. Strengthening Accountability of Actors Involved in Artificial Intelligence Activities

An AI system usually involves many actors: developers, providers, deployers, distributors, users, and third parties supplying data or infrastructure. Without clearly defined prohibited conduct, responsibility can easily be diffused and obscured behind claims that the harm was caused "by the algorithm" or "by the automated system." The prohibitions in Article 7 help arrest this tendency by identifying types of conduct for which any actor who performs, or knowingly participates in, such conduct must bear responsibility. The preventive value of accountability lies in forcing actors to adopt a compliance-by-design mindset from the outset.

4.4. Helping Shape a Responsible Model of Artificial Intelligence Development

Clearly identifying prohibited acts is not intended to impede innovation, but to shape a responsible model of AI development. An AI market that lacks clear legal limits may grow rapidly in the short term, but it also erodes social trust and increases long-term enforcement and remediation costs. By contrast, a legal framework capable of distinguishing between acceptable risk, controllable risk, and unacceptable risk creates conditions for sustainable innovation.

5. RECOMMENDATIONS FOR IMPROVING THE LAW AND ENHANCING ENFORCEMENT EFFECTIVENESS

First, guiding instruments should be issued promptly to clarify the open-textured concepts in Article 7, especially phrases such as "systematic manipulation of cognition," "serious harm," "exploitation of vulnerabilities," "distorting mechanisms of human control," and "information that must be transparent." If these concepts are not explained through concrete application criteria, the provisions may either be applied too narrowly and lose their preventive effect, or be applied too broadly and create instability for the innovation environment [3], [7], [8].

Second, a pre-deployment impact assessment mechanism should be added for AI systems

capable of significantly affecting human rights, the public interest, or data security. Such assessment should focus on the purpose of use, the affected groups, the types of data processed, the risks of bias, deception, or manipulation, the level of explainability and human oversight, and the incident-response plan. This approach is consistent with international trends in lifecycle governance and risk governance from the design stage [3], [4], [6].

Third, the law should more clearly codify data-governance obligations for AI activities. It is not enough to state generally that data must be "lawful"; the law should specify at least the purpose of processing, the legal basis, the categories of data sources, the principle of data minimization, retention periods, mechanisms for receiving and processing requests from data subjects, and the obligation to keep records of changes in training datasets. Practice at the EDPB and CNIL shows that legal disputes in AI increasingly revolve around the input-data chain rather than only the output product [9], [10].

Fourth, a minimum set of transparency obligations should be established for content generated or manipulated by AI. These obligations should include notifying users when they are interacting with an AI system; labeling deepfakes or synthetic content that may mislead; requiring providers to offer an appropriate level of provenance and traceability; retaining metadata or machine-readable markers; and establishing complaint mechanisms through which individuals may request verification, correction, or removal of content [7], [8].

Fifth, a mechanism for AI incident reporting and inter-agency enforcement coordination should be established among technology regulators, data authorities, cybersecurity bodies, consumer protection agencies, sector-specific regulators, and judicial authorities. Without an incident reporting system, regulators will lack empirical data to update policy; without inter-agency coordination, many violations will fall into jurisdictional gaps. The OECD currently regards monitoring, reporting, and learning from AI incidents as an important element of responsible risk governance [5].

Sixth, the compliance capacity of enterprises and the legal awareness of society should be strengthened. Enterprises should be guided in developing internal AI policies, data-audit procedures, human-control mechanisms, safety assessment files, and incident-response processes. At the same time, the public should be equipped with the skills needed to identify deepfakes, voice cloning, fraudulent chatbots, and other forms of AI-enabled behavioral manipulation. This is essential if the prohibitions are to function in practice rather than remain only on paper.

6. CONCLUSION

Prohibited acts in artificial intelligence activities are a central component of the 2025 Law on Artificial Intelligence, because they directly define the legal boundaries for highly dangerous AI practices. The greatest value of Article 7 lies not only in prohibiting individual acts, but in establishing "red lines" to protect human rights, maintain public trust, ensure accountability, and prevent systemic risks in the digital environment.

An analysis of each category of prohibited conduct shows that the law's underlying logic is to place technology within the limits of human dignity, autonomy, transparency, and the public interest. This is also a direction that is compatible with the AI governance standards currently emerging around the world. However, for these prohibitions to be genuinely effective, Vietnam must continue improving its implementing guidance, strengthening data governance and transparency for AI content, establishing pre-deployment impact assessments, building an AI incident reporting system, and enhancing compliance capacity in both the public and private sectors.

7. ACKNOWLEDGMENTS

I would like to thank TNU- University of Sciences for supporting us to complete this article.

REFERENCES

[1] National Assembly of Vietnam, Law No. 134/2025/QH15 on Artificial Intelligence, adopted 10 December 2025, effective 1 March 2026.

[2] National Assembly Web Portal, "National Assembly passes the Law on Artificial Intelligence: Regulating only AI outputs and patterns of use without hindering innovation," 10 December 2025.

[3] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024

laying down harmonised rules on artificial intelligence (Artificial Intelligence Act).

[4] Council of Europe, Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, opened for signature on 5 September 2024.

[5] OECD, AI risks and incidents.

[6] OECD Legal Instruments, Recommendation of the Council on Artificial Intelligence (OECD/LEGAL/0449), adopted 22 May 2019, amended 3 May 2024.

[7] European Commission, Code of Practice on marking and labelling of AI-generated content (policy page; drafting process under Article 50 AI Act).

[8] European Commission, Working Groups advance discussions on transparency obligations under Article 50 of the AI Act, 24 February 2026.

[9] European Data Protection Board (EDPB), Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models, 18 December 2024.

[10] CNIL, AI system development: CNIL's recommendations to comply with the GDPR, 5 January 2026.

[11] Federal Communications Commission, FCC 24-17, Implications of Artificial Intelligence Technologies on Protecting Consumers from Unwanted Robocalls and Robotexts, Declaratory Ruling, adopted 2 February 2024, released 8 February 2024; Federal Trade Commission, Approaches to Address AI-enabled Voice Cloning, 8 April 2024.

[12] U.S. Securities and Exchange Commission, SEC Charges Two Investment Advisers with Making False and Misleading Statements About Their Use of Artificial Intelligence, Press Release 2024-36, 18 March 2024.