The Structural Dilemma of AI-Driven Cybercrime
——The Dilution of Responsibility

Jibing Liu[1,a], Ran Xu [2*]

[1]College of Liberal Arts and Social Sciences, City University of Hong Kong, China
[2]Faculty of Law, The University of Macau, China

[a]liujliu2-c@my.cityu.edu.hk; *yc47237@edu.um.mo

Abstract. AI-driven cybercrime has led to a dilution of criminal responsibility along the chain of attribution. This dilution manifests at three levels: the spatio-temporal separation between conduct and consequence at the causal level; the opacity of causal chains at the subjective level; and the pluralization of responsible subjects at the attributable level. The traditional framework of criminal liability, grounded in the subjective element, objective element, and subject qualification, faces significant evidentiary challenges in the context of AI-driven cybercrime. Mainstream criminological theories have not entirely lost their explanatory power; however, certain underlying assumptions require targeted adjustments. These include the deterrence target presupposed by rational choice theory, and the guardianship capacity embedded in routine activity theory. This article proposes an analytical framework centered on the "dilution of responsibility" and, on this basis, advances a multi-tiered governance scheme. At the level of criminal law, it suggests introducing an algorithmic product liability regime modeled on product liability doctrine. At the administrative law level, it advocates for the establishment of mandatory security assessment mechanisms. In civil law, it proposes no-fault liability remedies for victims. At the level of social governance, it emphasizes AI literacy as a foundational means of reducing victim vulnerability. The theoretical contribution of this study lies in conceptualizing the impact of artificial intelligence on cybercrime as a structural dilution of responsibility rather than a fundamental transformation of the nature of crime, thereby providing constructive insights for both theoretical analysis and institutional design.

Keywords: Artificial Intelligence; Cybercrime; Criminal Responsibility; AI Literacy; Governance Mechanisms

1 Introduction

The traditional theory of criminal law is built upon a relatively clear causal chain. The perpetrator carries out a specific act, which directly or indirectly leads to a harmful consequence. The perpetrator is subjectively aware of the consequence, either intentionally or negligently, and therefore should bear criminal responsibility. The smooth operation of this attribution logic relies on several basic conditions: the traceability of

D. Magni et al. (eds.), *Proceedings of the 2026 3rd International Conference on Applied Economics, Management Science and Social Development (AEMSS 2026)*, Advances in Economics, Business and Management Research 389,
https://doi.org/10.2991/978-94-6239-672-2_67

the act and consequence in time and space, the recognizability of the causal chain, and the unity or divisibility of the responsible party. The deep involvement of artificial intelligence (AI) technology has not fundamentally overturned these conditions, but it has made it extremely difficult to satisfy them in practice.

Existing academic research on the relationship between AI and cybercrime can generally be divided into three categories. The first category focuses on how AI empowers cybercrime and discusses its dual nature, namely how AI can be used for prevention[1][2][3]. The second category, based on criminal law, explores how to incorporate AI-related actions into the existing criminal offense framework[4][5]. The third category, from a criminological perspective, analyzes how AI changes the structure of criminal motives and the distribution of opportunities[6][7]. These three types of research each have their value. Based on the above research, this article argues that the main challenge brought about by AI is not a qualitative change in the nature of the crime, but rather a quantitative accumulation of responsibility dilution in the attribution chain. It is this accumulation of quantitative changes that leads to difficulties in the practical application of the traditional attribution mechanism.

Based on this judgment, the research questions of this article can be precisely stated as follows: through what mechanisms does responsibility dilution occur in AI-driven cybercrime? What challenges does this dilution of responsibility pose to criminological theoretical frameworks and the criminal law attribution system? Can the existing legal system respond by making appropriate adjustments, or is it necessary to introduce new laws? What role should AI literacy occupy within the governance system? The answers to these questions constitute the main content of this article.

2 Three Levels of Responsibility Dilution

On the causal level, the dilution of responsibility arises from the extreme elongation of the temporal and spatial distance between the act and the resulting consequence. In traditional crimes, the will of the perpetrator and the execution of the act highly coincide in time and space. A thief's will directly controls the physical action at the moment of theft. Even in crimes involving tools, the use of those tools remains under the direct control of the perpetrator in real time. However, once a malicious AI program is deployed, it can autonomously execute attack behaviors months or even years later, based on changes in the environment, without any further involvement from the developer. While this spatial and temporal separation does not sever the causal relationship, it makes proving that relationship legally extremely complicated[8]. Demonstrating the criminal law-related causal link between an action from years ago and the specific harm today is a practical challenge in evidentiary law.

On the subjective level, the dilution of responsibility comes from the technical opacity of the causal chain. Modern AI systems, particularly those based on deep neural networks, have a well-known “black box” characteristic in their decision-making processes, meaning that even the system’s developers cannot fully explain how a specific input generates a specific output[9]. The legal consequence of this is that when it is necessary to prove that a subject “knew” or “should have foreseen” that the system

would produce a harmful consequence, a fundamental cognitive barrier arises. If even technical experts cannot fully foresee the system's behavior, how can the standard of "should have foreseen" be defined? The difficulty here lies in its objective limitation of technical cognition. Criminal law's subjective attribution, whether intentional or negligent, presupposes that the perpetrator has a certain level of cognitive ability, but when the object of cognition itself is technically opaque, this presumption becomes invalid.

On the attributable level, the dilution of responsibility is manifested in the multi-subject nature of the responsible parties and the ambiguity of responsibility shares. A complete AI-driven crime chain typically involves the developers of the underlying algorithms, the developers of the application tools, the platform service providers, and the final users. These subjects typically do not have the kind of conspiracy as seen in traditional accomplice crimes but instead are involved in commercial relationships formed through market transactions. The contribution of each subject to the final harmful consequence is fundamentally different in nature. The developers provided the technological possibility, the developers turned it into a tool, the platform provided the operating environment, and the users gave the specific criminal intent[10]. The current criminal law's theory of joint crime requires common intent or at least an understanding between the accomplices. However, in the above situation, each subject may be completely unaware of the final intent of others and may even be in entirely legal commercial relationships.

These three levels of dilution mechanisms have mutually reinforcing characteristics. The temporal and spatial separation makes the factual determination of causality more reliant on technical verification, while the technical opacity casts doubt on the reliability of the verification itself. The multi-subject structure means that even if the causal relationship is established, the division of responsibility remains challenging. The overlap of these factors results in obstacles at every stage of criminal prosecution, including evidence collection, qualification, and sentencing. However, this dilution of responsibility is not absolute. In some cases, traditional methods can still successfully hold individuals accountable, but it is structural, meaning it significantly reduces the probability of successful prosecution in statistical terms. It is this structural reduction in probability, rather than the absolute impossibility of prosecution in individual cases, that presents the true challenge to governance of cybercrime in the AI era.

3 Appropriate Explanation and Adjustments to Criminological Theory

The task of criminology is to explain why crimes occur. When assessing the impact of AI on criminological theory, two tendencies must be avoided. One is being overly pessimistic, assuming that the existing theoretical framework has completely failed. The other is being overly optimistic, thinking that simply adding AI to the list of variables will suffice. The accurate judgment is that the core logic of mainstream criminological theories remains valid, but the specific assumptions upon which their operation depends are challenged by AI, requiring targeted adjustments to the theory.

Rational Choice Theory views crime as a rational decision made by the perpetrator after weighing the benefits and costs, meaning that deterrence can be achieved by increasing the cost of crime[11]. The issue AI-driven cybercrime faces in this theory is the misalignment of the deterrence target. In traditional crime, The target of deterrence in punishment is closely aligned with the executor of the crime. In AI-driven cybercrime, the real executor of the harmful act is the algorithmic system, and AI systems lack the ability to perceive the deterrence effect of punishment. While punishment can theoretically be traced backward to deter the developers or deployers of the system, there is a challenge in transmitting the deterrence effect[8]. When a developer knows that their tool may be misused for crime but the marginal cost of developing it is extremely low, and once the tool is in circulation, it becomes difficult to control, deterring the developer doesn't effectively transmit the deterrence to the large number of potential tool users. Therefore, the adjustment direction for rational choice theory is not to abandon deterrence but to recognize that deterrence must shift upstream to the technological chain and must be combined with technological control measures to be effective.

Routine Activity Theory suggests that crime occurs when three elements converge in time and space: a motivated offender, a suitable target, and the absence of capable guardianship[12]. This theory still holds explanatory power for AI-driven crime, but the meaning of capable guardianship has changed significantly. Traditional guardianship primarily relied on the presence of people in physical space, such as neighbors'attention, security patrols, and the deterrence of surveillance equipment. In cyberspace, guardianship heavily depends on technological capabilities, such as firewalls, intrusion detection systems, and anomaly detection algorithms[13]. AI-driven attack systems exponentially increase in speed, scale, and complexity, causing traditional technological defenses to lag behind, creating a generational gap in capability[14]. The adjustment to routine activity theory, therefore, is not in modifying the theory itself but in recognizing that effective guardianship in the technological age must include the continuous upgrading of technological capabilities, and this upgrading needs institutional investment to ensure it.

Overall, we do not need to completely discard existing criminological theories, but we must, when applying these theories, clearly identify which assumptions need to be adjusted under new technological conditions and which policy tools need targeted improvements.

4 The Proof Dilemma of Criminal Law Attribution Requirements

The proof dilemma for subjective elements is primarily reflected in the blurred standards for determining intent and negligence. Chinese criminal law requires that intent be defined as the perpetrator "knowing" acting with awareness of the nature and consequences of their behavior, while negligence is defined as the perpetrator "should have foreseen" the consequences but failed to do so. In the context of AI-driven cybercrime, the standards for "knowing" and "should have foreseen" are both vague. When a developer releases an AI tool with potential criminal uses, can it be determined that they

knew the tool would be used for a crime? This depends on the relationship between the tool's design purpose, functional characteristics, and actual usage scenario. A general-purpose tool and a specialized tool (such as an automated phishing email generator) should be distinguished in terms of the determination of "knowing." But where are the boundaries of this distinction? How does the law define the significance of potential criminal use? This requires the support of technical expertise, and there are disputes over the standards of technical evaluation itself.

The determination of negligence is even more complex. Negligence theory assumes that the perpetrator has the ability to foresee specific dangers, but when the danger arises from the complex interactions of a technological system, how should the standard for "should have foreseen" be determined? One possible approach is to draw from product liability law's reasonable care standard, which includes industry-accepted safety measures, foreseeable misuse scenarios, and affordable preventive measures. However, the application of this standard depends on the maturity and stability of industry standards, and AI technology is still in a rapid iteration phase, with industry standards yet to be fully developed. In such cases, how should "reasonable care" be defined, and does the court have the capacity to judge the sufficiency of technical safety measures in individual cases? These are practical issues that urgently need to be addressed.

The proof dilemma for objective elements is primarily reflected in the determination of causality. Criminal law requires proving "without the preceding act, there would be no subsequent result." In AI-driven cybercrimes, the complexity of the causal chain creates obstacles for this proof. A typical scenario is when an AI system, after deployment, develops functions not originally included in its design through autonomous learning, and these functions are exploited to commit a crime[15]. In this case, how is the causal relationship between the deployment act and the criminal result determined? If the condition theory is applied, the deployment act is clearly a necessary condition. However, if the proximate cause theory is used, it must be proven that the system's autonomous evolution is sufficiently proximate, which may be technically unpredictable, thus failing to meet the standard of proximity. The choice of the causal relationship determination standard directly affects the severity of the attribution, but legal theory has not yet reached a consensus on this matter.

The difficulty with subjective elements lies in the distribution of responsibility in multi-party situations. As mentioned earlier, the AI-driven cybercrime chain involves multiple parties, and their contributions to the harmful consequences differ fundamentally in nature. The current criminal law's theory of joint crime centers on shared intent. However, in the technological chain, there is often no traditional conspiracy between the parties. One possible legal path is to apply the theory of indirect perpetration, treating the downstream user as a tool and holding the upstream developer accountable. But this approach applies only when the downstream user lacks criminal responsibility or is deceived, whereas in AI-driven cybercrime, the downstream user is typically a fully responsible intentional criminal, making it difficult to include them in the indirect perpetration framework[5]. Another approach is to expand the application of accessory liability, treating the provision of criminal tools as an act of assistance. However, for accessory liability to apply, the accomplice must know of the other's criminal intent, and

the tool developer faces an unspecified market, making it challenging to prove their "knowing." This, again, returns to the proof dilemma for subjective elements.

In response to the above dilemmas, the direction for criminal law reform should be fine-tuned adjustments. On the subjective element level, a reasonable expectation standard from product liability law can be used to establish a duty for AI system developers to foresee misuse scenarios, with a failure to fulfill this duty serving as the basis for determining negligence. On the causality level, the proof standard can be moderately lowered in specific types of AI-driven cybercrimes, allowing causality to be established as long as the defendant's actions significantly increased the risk of harm occurring. On the subject element level, a joint liability mechanism similar to that in product liability law could be created. For high-risk AI products, developers, platforms, and users would bear joint liability for damages to victims, with internal recourse based on the degree of fault[16]. The common feature of these adjustments is that, while retaining the basic principles of criminal law, they enhance the operability of legal application through moderate adjustments to the proof of responsibility and attribution standards.

5 Sources of Victim Vulnerability and AI Literacy

In AI-driven cybercrime, the cognitive abilities of victims often become the key factor in determining whether the crime succeeds. The success of a crime largely depends on the criminal's ability to exploit the victim's cognitive weaknesses. For example, deepfake technology deceives people's sensory perceptions, algorithmic recommendation systems manipulate the information environment, and automated social engineering attacks precisely target a person's psychological defenses[17].

The vulnerability of victims primarily arises from a generational cognitive gap. The development speed of AI technology far exceeds the pace at which the general public updates its understanding of new technologies. Five years ago, most people had never heard of deepfakes. Today, anyone with access to generative AI apps can create convincing fake audio and video content. However, public awareness of what technology is capable of still lags far behind the knowledge of those who control the technology. There is a significant information gap between technology controllers and ordinary users, with the former always being aware of the technology's potential before the latter. When criminals use technological methods that the public has not yet recognized, even the most vigilant victims are unlikely to defend themselves, because you cannot defend against a threat you do not even know exists.

Victim vulnerability is also closely tied to social inequality. Research shows that differences in digital literacy are often closely related to socioeconomic status, education level, and age. Elderly people, those with low educational levels, and low-income users have significantly higher victimization rates in AI-driven cybercrime[18]. This disparity should be viewed as the reproduction of inherent structural inequalities in the digital age. These groups are less likely to encounter digital technology in their daily lives, lack tacit knowledge accumulated through trial and error, and thus struggle to develop an instinctive alertness when faced with cybercrime. More concerning is the

fact that AI-driven crime tools can use algorithmic analysis to accurately identify and prioritize attacking these cognitively vulnerable groups. This means that the risk of victimization is not randomly distributed but is structurally concentrated in vulnerable groups[19].

AI literacy should encompass three aspects. The first aspect is the ability to recognize, which is the ability to judge the authenticity of information. Large language models can generate fluent, coherent, and emotionally rich text. Deepfake technology can synthesize hyper-realistic faces, voices, and videos. Recommendation algorithms, before users even notice it, shape the boundaries of their information reception through echo chambers and filter bubbles[20]. In this technological context, individuals who lack recognition skills are highly susceptible to accepting false information unknowingly, being manipulated by inflammatory content, or mistakenly treating AI-generated outputs as authoritative in critical decision-making. Previously, people's ability to recognize information relied on intuition and experience to distinguish truth from falsehood. However, today, recognition ability cannot be limited to this; it must be based on an understanding of the basic principles of AI. People need to understand what training data is, what hallucination is, and how recommendation algorithms build user profiles based on behavioral data. Only by understanding what AI can do can individuals detect what it has done in specific situations. In this sense, recognition ability is also a manifestation of critical thinking skills.

The second aspect is the ability to use, which refers to whether an individual can effectively and appropriately apply AI tools to real tasks. This dimension is often simplified to operational proficiency, such as being able to construct effective prompts to get generative AI tools like ChatGPT to write copy or generate images. However, the ability to use AI has a deeper meaning. First, the ability to use is reflected in situational judgment. An individual with true use capability can assess whether it is appropriate to introduce AI tools into a specific context and to what extent AI participation should be limited. In scenarios such as decision support in medical diagnoses, risk assessments in judicial sentencing, and automatic grading in educational evaluations, there is a risk of simplifying complex value judgments into algorithmic outputs. A person with use capability can recognize this risk and make a conscious decision. Second, the ability to use involves the maintenance of subjectivity. The convenience of AI tools easily induces cognitive laziness[21]. When AI can quickly generate an analysis report or a creative proposal, will individuals still retain the willingness and ability to think independently? If the vast majority of people delegate thinking to AI, the consequence will be an irreversible collective intellectual decline. Therefore, true use capability is using AI to expand the depth and breadth of one's own thinking, while always maintaining the initiative to critically assess, select, and recreate the output results. Only those who can do this can be considered as having the ability to use AI.

The third aspect is the ability to protect, which refers to an individual's capacity to effectively protect their own interests and avoid harm from technical systems during interactions with AI. This ability is rooted in the inherent asymmetry of AI systems. Users are often unaware of what data the system collects or what logic it operates on, while developers and platforms have access to all the information[22]. Protective ability

covers multiple levels. On the data level, individuals need to understand how AI systems collect and use personal data, and be able to reasonably exercise privacy settings, data authorization reviews, and other basic operations. On the content level, individuals need to master methods of information verification and cross-checking. When engaging with AI-generated content, they should have the proactive awareness to trace sources and discern authenticity. On the psychological level, individuals need to remain vigilant about the information echo chambers and emotional content created by recommendation algorithms, recognize the commercialization of attention and emotions by platforms, and proactively build diverse information intake channels.

However, the most fundamental challenge facing protective ability is that the risks it addresses are often intangible, and difficult to directly perceive[23]. Algorithmic bias cannot be clearly identified like a discriminatory statement. The continuous accumulation and commercial use of data is also not presented to users in the form of notifications. The formation of information echo chambers is an even more hidden process. This invisibility means that protective ability cannot be limited to responding to specific threats, but must develop into a long-term capacity for self-observation and reflection, to examine how technology shapes an individual's cognition and behavior.

6 Conclusion: Multi-Layered Governance

The analysis above indicates that the governance of AI-driven cybercrime involves not only the difficulties of criminal prosecution, but also the insufficient capacity for crime prevention and the lack of victim protection systems, requiring collaborative efforts across multiple levels.

On the criminal law level, a legal mechanism for algorithmic product liability can be introduced. This mechanism is analogous to product liability law but applies to AI products that have a significant risk of criminal misuse. Developers or operators who knowingly or should have known that their products present a significant risk of criminal misuse, and who fail to take reasonable preventive measures, should be held criminally liable for any serious consequences resulting from the use of the product in a crime. The key to this mechanism is to clarify three elements: First, the standard for determining significant criminal misuse risk, which can refer to risk assessment techniques and the establishment of a tiered classification system. Second, the definition of reasonable preventive measures, which should include technical means and management measures such as user identity verification, usage scenario limitations, and suspicious behavior monitoring. Third, the form of liability can be based on negligent criminal conduct, with sentencing linked to the actual harm caused.

On the administrative law level, a mandatory security assessment and filing system for AI systems should be established. This system targets high-risk application scenarios such as financial services, medical diagnostics, and autonomous driving. The core requirement is that, before the system is put into use, it must undergo an independent third-party security assessment, which should evaluate technical security, data compliance, potential social risks, and more. Once the assessment is passed, the system can be filed with the regulatory authorities and put into use. During operation, the system

should be continuously monitored and report any anomalies regularly. The key to implementing this is building the professional capabilities of third-party assessment institutions and enhancing the technical oversight capabilities of regulatory agencies.

On the civil law level, a no-fault relief mechanism should be established for victims of AI-driven crimes. The difficulties of criminal prosecution should not translate into obstacles for victim relief. Drawing from the design of mandatory liability insurance in traffic accidents, it can be required that developers or operators of high-risk AI products purchase liability insurance. When the product is used in a crime causing harm, victims can directly claim compensation from the insurance company without needing to prove the fault of the developer or operator. The insurance mechanism allows for the transfer of significant individual compensation cases into an industry-wide risk-sharing model, ensuring victim protection while preventing burdensome individual liabilities from stifling technological innovation.

On the social governance level, the cultivation of AI literacy should be promoted as a long-term, foundational project. In education, digital literacy should be incorporated into basic education curricula, with a focus on developing students'critical thinking skills. In public information, a timely early warning mechanism for new types of technological fraud should be established, with risk alerts pushed to the public through mainstream media and social platforms. Additionally, technology companies should be encouraged to take on social responsibility by embedding safety risk alerts functions in product designs, reducing the cognitive burden on users. In legal rights, the “right to explanation” for citizens in relation to AI systems should be clearly defined, and convenient channels for exercising this right should be established.

These four layers of governance form a complete system for responsibility allocation and risk control. Criminal law deters developers in advance through criminal responsibility. Administrative law implements process supervision through mandatory assessments and filings. Civil law ensures post-crime relief through no-fault liability insurance. Social governance reduces victim vulnerability through the cultivation of AI literacy and risk alerts. The collaboration of these four layers collectively points toward the same goal: under the condition of rapid technological evolution, to minimize the space for responsibility dilution, ensure that legal attribution remains technically feasible, crime prevention is within reach, and victim relief is institutionally available.

References

1. Blauth, T. F., Gstrein, O. J., & Zwitter, A. (2022). Artificial Intelligence Crime: An Overview of Malicious Use and Abuse of AI. *IEEE Access, 10,* 77110–77122. https://doi.org/10.1109/access.2022.3191790.
2. Kaur, R., Gabrijelčič, D., & Klobučar, T. (2023). Artificial intelligence for cybersecurity: Literature review and future research directions. *Information Fusion, 97*(101804), 1–29. ScienceDirect. https://doi.org/10.1016/j.inffus.2023.101804
3. Amir Djenna, Barka, E., Achouak Benchikh, & Khadir, K. (2023). Unmasking Cybercrime with Artificial-Intelligence-Driven Cybersecurity Analytics. *Sensors, 23*(14), 6302–6302. https://doi.org/10.3390/s23146302

4. Sachoulidou, A. (2024). AI Systems and Criminal Liability: A Call for Action. *Oslo Law Review, 11*(1), 1–10. https://doi.org/10.18261/olr.11.1.3
5. Panattoni, B. (2025). Generative AI and criminal law. *Cambridge Forum on AI: Law and Governance, 1*. https://doi.org/10.1017/cfl.2024.9
6. Hayward, K. J., & Maas, M. M. (2020). Artificial Intelligence and Crime: A Primer for Cri minologists. Crime, Media, *Culture: An International Journal, 17*(2). https://doi.org/10.11 77/1741659020917434
7. Tong, M. (2025). The New Actor: Artificial Intelligence in Criminology and Criminal Justice. Journal of Criminology, *Criminal Justice, Law & Society, 26*(2), 12–22. https://doi.org/10.54555/ccjls.13020.146572
8. Nerantzi, E., & Sartor, G. (2024). "Hard AI Crime": The Deterrence Turn. *Oxford Journal of Legal Studies, 44*(3). https://doi.org/10.1093/ojls/gqae018
9. Yavar Bathaee. (2018). The Artificial Intelligence Black Box and the Failure of Intent and Causation. *Harvard Journal of Law & Technology, 31*(2), 889.
10. Maskur, M. A., Masyhar, A., Damayanti, R., Ramada, D. P., & Sanyal, S. (2025). Reimagining Criminal Liability in the Age of Artificial Intelligence: Toward a Comparative and Reform-Oriented Legal Framework. *Journal of Law and Legal Reform, 6*(4), 1805–1838. https://doi.org/10.15294/jllr.v6i4.35540
11. Cornish, D. B., & Clarke, R. V. (1986). The reasoning criminal: Rational choice perspectives on offending.
12. Cohen, L. E., & Felson, M. (1979). Social change and crime rate trends: A routine activity approach. *American Sociological Review, 44*(4), 588–608. https://doi.org/10.2307/2094589
13. Williams, M. L. (2016). Guardians Upon High: An Application of Routine Activities Theory to Online Identity Theft in Europe at the Country and Individual Level. *British Journal of Criminology, 56*(1), 21–48. https://doi.org/10.1093/bjc/azv011
14. Nespoli, P., Molina, S. B., Beltrán-López, P., & Mármol, F. G. (2025). Tackling Cyberattacks Through AI-Based Reactive Systems: A Holistic Review and Future Vision. *IEEE Communications Surveys & Tutorials*.
15. King, T. C., Aggarwal, N., Taddeo, M., & Floridi, L. (2020). Artificial intelligence crime: An interdisciplinary analysis of foreseeable threats and solutions. *Science and Engineering Ethics, 26*(1), 89–120. https://doi.org/10.1007/s11948-018-00081-0
16. Bhatt, N. (2025). Crimes in the age of artificial intelligence: A hybrid approach to liability and security in the digital era. *Journal of Digital Technologies and Law, 3*(1), 65-88.
17. Schmitt, M., & Flechais, I. (2024). Digital deception: Generative artificial intelligence in social engineering and phishing. *Artificial Intelligence Review, 57*(12), 324. https://doi.org/10.2139/ssrn.4602790
18. Ragnedda, M., Ruiu, M. L., & Addeo, F. (2022). The self-reinforcing effect of digital and social exclusion: The inequality loop. *Telematics and Informatics, 72*, 101852.
19. Joyce, K., Smith-Doerr, L., Alegria, S., Bell, S., Cruz, T., Hoffman, S. G., Noble, S. U., & Shestakofsky, B. (2021). Toward a Sociology of Artificial Intelligence: A Call for Research on Inequalities and Structural Change. *Socius: Sociological Research for a Dynamic World, 7*, 1–11. https://doi.org/10.1177/2378023121999581
20. Ahmmad, M., Shahzad, K., Iqbal, A., & Latif, M. (2025). Trap of Social Media Algorithms: A Systematic Review of Research on Filter Bubbles, Echo Chambers, and Their Impact on Youth. *Societies, 15*(11), 301. https://doi.org/10.3390/soc15110301
21. Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Dragan Gašević. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology, 56*(2). https://doi.org/10.1111/bjet.13544

22. Felzmann, H., Villaronga, E. F., Lutz, C., & Tamò-Larrieux, A. (2019). Transparency You Can trust: Transparency Requirements for Artificial Intelligence between Legal Norms and Contextual Concerns. *Big Data & Society, 6*(1), 1–14. https://doi.org/10.1177/2053951719860542
23. Diberardino, N., Baleshta, C., & Stark, L. (2024). Algorithmic harms and algorithmic wrongs. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency* (pp. 1725-1732).