IRSHAD
Uzbek Journal of Law and Digital Policy |
Volume: 3, Issue: 6
2025

# Challenges in Regulating and Prosecuting AI Model Poisoning as Cybercrime

Naeem AllahRakha
Tashkent State University of Law

## Abstract

This research examines the critical challenges confronting legal systems in regulating and prosecuting AI model poisoning as cybercrime. Through qualitative doctrinal analysis and comprehensive document review, the study evaluates international legal frameworks addressing AI poisoning, explores prosecution difficulties including proving intent and establishing liability, and assesses regulatory roles in preventing incidents. Findings reveal significant gaps in existing cybercrime statutes that fail to recognize AI poisoning as distinct offenses, creating uncertainty for law enforcement. The automated nature of machine learning obscures causation chains, making liability determinations nearly impossible under traditional legal principles. Cross-border enforcement fails because international agreements like the Budapest Convention lack specific provisions for AI attacks spanning multiple jurisdictions. Courts operate without precedents, forcing reliance on inadequate analogies to conventional cybercrimes. The research recommends enacting comprehensive legislation explicitly criminalizing AI poisoning, updating international treaties to facilitate cooperation, establishing mandatory security standards for high-risk systems, and developing specialized forensic capabilities within law enforcement agencies to address these emerging technological threats effectively.

**Keywords**: AI Model Poisoning, Cybercrimes, Prosecution, Legal Frameworks, Algorithmic Manipulation

**APA Citation:**

AllahRakha, N. (2025). Challenges in Regulating and Prosecuting AI Model Poisoning as Cybercrime. *Uzbek Journal of Law and Digital Policy, 3*(6), 28–51. https://doi.org/10.59022/ujldp.472

## I. Introduction

A poisoned dataset can silently turn trusted AI into a dangerous legal risk. This threat, known as AI model poisoning, involves corrupting training data or algorithms (Cotroneo et al., 2024). Unlike traditional cybercrime, the harm appears later and is often difficult to trace. This makes regulation and prosecution especially challenging for existing criminal law systems. Most of the laws were written before artificial intelligence became central to public and private decisions. As a result, key legal concepts like intent, harm, and causation become unclear. Jurisdiction problems also arise when poisoned models' cross borders through global digital networks. Proving responsibility is harder when attacks involve anonymous actors or automated processes.

AI systems increasingly influence policing, finance, healthcare, and other sensitive sectors worldwide. These systems depend on datasets, making them vulnerable to manipulation during training stages. AI poisoning emerged as attackers learned to corrupt data rather than attack software (Iyer, 2023). Early research focused on technical detection methods, leaving legal analysis limited and fragmented. Most cybercrime laws were designed for hacking, fraud, or data theft offenses. They rarely address indirect harms caused by altered models producing harmful outputs. Scholars debate whether existing criminal principles can cover hidden and delayed AI harms. Some studies suggest civil liability fits better, but criminal accountability remains uncertain. Cross border data flows further complicate enforcement, jurisdiction, and evidence collection processes.

Legal systems already recognize cybercrime, but mainly focus on direct and visible attacks. We know that poisoned models may act lawfully on the surface yet cause unlawful outcomes. Current laws have less capacity to classify all acts as crimes under traditional definitions. The main problem is the lack of clear legal standards for prosecuting AI poisoning. Prosecutors face difficulties proving intent, causation, and damage beyond reasonable doubt. Responsibility is unclear when harm results from shared data and complex development chains. We still lack guidance on applying criminal liability to automated and learning systems. There is also uncertainty about which actors should be legally accountable for poisoning acts. Cross border use of AI further weakens enforcement and evidence collection efforts.

Existing research on AI model poisoning highlights important technical and regulatory debates, but reveals clear legal gaps needing deeper study. AI's rapid development creates new forms of cybercrime that strain current legal systems, including criminal definitions, evidence standards, and enforcement practices (Sun et al., 2026). Despite strong technical analysis of data poisoning risks and defensive techniques, most work focuses on detection methods rather than legal accountability or prosecution (Zhang et al., 2025). Literature on AI and criminal law more broadly discusses liability challenges around mens rea and actus reus, but without specific guidance on model

poisoning incidents (Panattoni, 2025). A major weakness is the lack of empirical or doctrinal analysis on how to classify and prosecute AI poisoning under existing cybercrime statutes. This gap suggests future research should develop concrete legal frameworks and prosecutorial standards for AI poisoning as an identifiable cybercrime.

Despite growing research on AI security and cybercrime, significant legal gaps remain regarding AI model poisoning. Current studies focus mainly on technical detection and mitigation strategies, leaving prosecution and legal accountability underexplored. The challenges in proving intent, assigning responsibility, and establishing causation in AI poisoning cases, but offers limited solutions. Most analyses are theoretical, with few empirical studies examining how courts or regulators handle such incidents. Furthermore, cross-border enforcement and jurisdictional issues are often noted but not systematically studied, creating uncertainty for global AI applications. The gap indicates a need for research that develops clear legal frameworks, prosecutorial guidelines, and policy recommendations to classify and address AI model poisoning effectively as a cybercrime. The researches objectives for your study:

To evaluate international legal frameworks and approaches in addressing AI model poisoning.

To examine difficulties in prosecuting AI poisoning cases, including proving intent, causation, and liability.

To assess the role of regulations in preventing and responding to AI poisoning incidents.

***What are the key legal challenges in regulating and prosecuting AI model poisoning as a form of cybercrime, and how can existing laws be adapted to ensure accountability and effective enforcement?***

AI model poisoning is a silent cyber threat that could destabilize critical systems before anyone even notices. Today's laws are struggling to keep up, leaving gaps that allow malicious actors to exploit AI with little fear of accountability. This research is urgent because it exposes these legal blind spots and explores how regulation and prosecution can catch up to technology. Academically, it fills a critical void in understanding AI-related cybercrime and offers a roadmap for future legal scholarship. Practically, it provides policymakers, prosecutors, and regulators with concrete strategies to identify, prevent, and punish AI poisoning. Societally, it protects people, organizations, and public trust in AI systems, ensuring that innovation does not outpace justice. By confronting the invisible dangers of AI, this study delivers timely, actionable insights that can shape law, policy, and the safe future of technology.

## II. Methodology

This study uses a qualitative research design to explore the legal challenges in regulating and prosecuting AI model poisoning as a cybercrime. Qualitative methods are suitable because the research focuses on analyzing laws, regulations, and scholarly

literature rather than numerical data. This approach allows a detailed understanding of legal frameworks, gaps, and challenges in applying criminal law to AI-related threats. The target population includes existing laws, regulations, and legal frameworks on cybercrime and AI-related offenses, as well as scholarly articles discussing AI model poisoning, liability, and enforcement. The sample consists of selected legal documents accessed through official government portals and relevant peer-reviewed journal articles retrieved using specific keywords such as "AI poisoning," "cybercrime law," "legal accountability AI," and "AI regulation." The selection criteria include relevance, credibility, and publication within the last five years.

Data were collected through systematic searches on Google Scholar using the selected keywords and from official legal portals for statutes, regulations, and policy documents. All sources were publicly available to ensure transparency and accessibility. The study relies on publicly available legal documents and scholarly literature. Legal instruments, such as data protection laws, cybercrime statutes, and regulatory guidelines, were retrieved from official government websites. Scholarly articles were sourced from peer-reviewed databases via Google Scholar, ensuring credibility and relevance. To ensure validity, only peer-reviewed articles and official legal documents were included. The literature is limited to publications within the last five years to maintain currency. Reliability is ensured by cross-verifying legal provisions and scholarly arguments across multiple sources. All sources are cited accurately to acknowledge the original authors and maintain research integrity.

A doctrinal analysis approach was used. Legal documents were examined to identify principles, definitions, and enforcement mechanisms relevant to AI model poisoning. The study used only publicly available data and did not involve human participants, ensuring minimal ethical risks. There was no conflict of interest, and the study was conducted solely for academic research purposes. The study is limited to legal documents and scholarly literature from the last five years and focuses on English-language sources. The research relies on publicly available documents, which may not capture confidential enforcement practices or unpublished case studies. It is assumed that the selected legal documents and scholarly literature are accurate, credible, and representative of the broader legal and academic discussions on AI model poisoning.

## III. Results

This study examined the legal challenges in regulating and prosecuting AI model poisoning as a cybercrime. Using qualitative analysis of scholarly articles and legal documents, the research explored key gaps in laws, enforcement mechanisms, and accountability frameworks. The research focused on understanding how current laws address AI-related harms, the difficulties prosecutors face in proving intent and causation, and the responsibilities of different actors involved in AI systems. The findings provide insight into both technical and legal dimensions of AI poisoning, highlighting

areas that require urgent attention from policymakers, regulators, and legal scholars.

| Category | Findings | Implications |
|---|---|---|
| **Legal Gaps in AI Cybercrime Laws** | Most cybercrime statutes target traditional hacking, fraud, and data theft. AI poisoning is not clearly classified as a crime. | Legal definitions need updating to include AI-specific harms. |
| **Challenges in Proving Liability** | Automated systems and multi-party data chains obscure intent and causation. Responsibility among developers, deployers, and attackers is unclear. | Clear liability frameworks are required. |
| **Cross-Border Enforcement Issues** | AI poisoning often involves multiple countries, creating jurisdictional challenges. International agreements provide limited guidance. | The need for global cooperation and coordinated legal frameworks. |
| **Lack of Case Law and Precedents** | Very few judicial cases exist on AI poisoning; courts rely on analogies to traditional cybercrime. | Legal interpretations are inconsistent. |

Analysis of cybercrime statutes across jurisdictions revealed that most existing laws are designed for traditional hacking, data theft, and fraud. They rarely account for harms caused indirectly through poisoned AI models (Sarkar & Shukla, 2023). Definitions of intent, causation, and damage are insufficient to classify AI poisoning as a criminal act. This gap creates uncertainty for law enforcement agencies and exposes society to risks from undetectable AI manipulation. The study found that establishing responsibility in AI poisoning cases is highly complex. Automated decision-making and multi-party data processes obscure the chain of causation, making it difficult to prove criminal intent or negligence. Legal frameworks lack clear guidance on how to assign accountability between developers, deployers, and attackers, leaving prosecutors without practical tools for enforcement.

AI poisoning incidents often involve data and systems spanning multiple countries, raising significant jurisdictional challenges. Existing international cybercrime agreements provide limited support for enforcement in such cases. This highlights the need for

coordinated global legal approaches and collaboration among regulatory authorities to effectively prosecute AI-related cybercrimes. The research identified very few reported legal cases specifically addressing AI poisoning. This absence of judicial precedents limits the ability of courts to interpret existing laws in this context. Scholars emphasize that without concrete case studies, prosecutors and judges must rely on analogies to traditional cybercrime, which may be insufficient to address AI-specific harms (Zaidan & Ibrahim, 2024).

The findings directly address the research question by showing that current legal frameworks are inadequate for regulating and prosecuting AI model poisoning. The study demonstrates that the main legal challenges involve defining AI poisoning as a crime, proving intent and causation, and allocating responsibility among multiple actors. It also highlights gaps in cross-border enforcement and the lack of judicial precedents. By identifying these challenges, the research provides a foundation for developing clearer legal definitions, prosecutorial guidelines, and international cooperation strategies, thereby addressing the objectives of understanding legal gaps, prosecution difficulties, and potential frameworks for accountability.

## IV. Discussion

### A. Inadequacy of Existing Cybercrime Laws to Address AI Model Poisoning

Most cybercrime laws were written before modern artificial intelligence became widespread. These laws mainly focus on acts like unauthorized access, data theft, fraud, or system interference. AI model poisoning does not fit easily into these categories. As a result, harmful actions that manipulate training data or model behavior often fall outside clear criminal definitions. This creates uncertainty for regulators, law enforcement, and courts. It also creates risk for society as AI systems are increasingly used in healthcare, finance, education, and public administration. AI model poisoning involves intentionally inserting harmful or misleading data into training sets (Allheeib, 2024). This can cause models to behave in unsafe or biased ways. In some cases, the harm is subtle and delayed. Unlike traditional hacking, there may be no system break-in. The attacker may use legitimate access points, open data sources, or shared platforms. Existing cybercrime laws usually require proof of unauthorized access or direct damage. When these elements are missing, prosecution becomes difficult.

Legal analysis from recent academic studies shows that many jurisdictions lack clear language addressing AI-specific harms. For example, the Computer Misuse Act in the United Kingdom focuses on access offenses. The United States Computer Fraud and Abuse Act has similar limits. Neither law clearly addresses manipulation of training data where access is lawful. This shows that the problem is structural rather than accidental. Laws were designed for a different technological era. The evidence supporting this finding comes from statutory reviews, expert commentary, and policy consultations. The significance of this finding lies in its impact on accountability. When harmful AI

behavior occurs, victims may struggle to seek justice. Law enforcement agencies face uncertainty about which charges to apply. Prosecutors may avoid cases due to low chances of success. This weakens deterrence. It also undermines public trust in AI technologies. If people believe that AI harms are legally invisible, acceptance of AI systems may decline. This has economic and social consequences.

Recent policy initiatives confirm the existence of these gaps. The European Union's Artificial Intelligence Act focuses mainly on risk management and compliance. It does not clearly criminalize AI poisoning. Instead, it relies on administrative penalties and oversight. While this is a step forward, it does not fully address intentional malicious acts. Similarly, the United Nations discussions on cybercrime have not yet produced AI-specific criminal provisions. The strength of the evidence lies in its consistency across regions. Legal scholars from different countries identify similar weaknesses. This suggests that the issue is global rather than local. However, there are also limitations. Much of the evidence is based on theoretical analysis rather than real cases. Because AI poisoning cases are rarely reported, it is hard to measure how often these gaps are exploited. This limits empirical certainty. Still, the absence of cases itself supports the argument that legal clarity is missing.

Potential bias may arise from the focus on formal law rather than informal enforcement. In practice, some AI harms may be addressed through civil liability, contract law, or regulatory sanctions. These responses are often excluded from cybercrime analysis. This may exaggerate the sense of a legal vacuum. However, civil remedies do not replace criminal accountability. They often require resources that victims lack. Therefore, the core concern remains valid. Comparisons with data protection law reveal partial overlap. For instance, poisoning that introduces biased data may violate fairness principles under the GDPR. Yet data protection law focuses on personal data, not model integrity. Many AI systems use non-personal data. This leaves large areas uncovered. Intellectual property law also offers limited help. It protects ownership, not safety or trust. These comparisons show that existing legal tools are fragmented and indirect.

Another influencing factor is the role of private companies. Many AI systems are developed and trained by large firms. They control data pipelines and model updates. When poisoning occurs, it may be unclear whether it was an external attack or an internal failure (Diro et al., 2025). Companies may be reluctant to report incidents due to reputational risk. This reduces visibility and slows legal development. It also shapes how laws are written, as policymakers often rely on industry input. Technological complexity also influences legal gaps. AI systems are difficult to explain. Legislators may hesitate to create criminal offenses they do not fully understand. This leads to broad principles rather than precise rules. While flexibility can be useful, it also creates uncertainty. Criminal law requires clarity. Without clear definitions of AI poisoning, intent, and harm,

enforcement remains weak.

In 2023 and 2024, several governments funded research on AI security and integrity. The United States National Institute of Standards and Technology released guidance on managing AI risks. These initiatives focus on prevention rather than punishment. They support the idea that regulation is currently more concerned with governance than crime. This aligns with the research finding that criminal law has not kept pace. Lawmakers need to update cybercrime laws to include AI-specific offenses. Definitions should cover manipulation of training data and models, even when access is authorized. Intent standards should reflect the realities of automated systems. Harm should include long-term and indirect effects. International cooperation is also needed to harmonize definitions. Without these changes, accountability will remain limited. In practical terms, this research applies to regulators, prosecutors, and developers. Regulators can use it to justify legal reform. Prosecutors can use it to argue for clearer mandates. Developers can use it to advocate for shared standards and reporting mechanisms.

### B. Difficulty in Establishing Criminal Liability within Complex AI Development Chains

The complexity of establishing liability in AI model poisoning cases represents perhaps the most formidable obstacle facing legal systems worldwide. Unlike traditional cybercrimes where perpetrators directly access systems and cause immediate, visible damage, AI poisoning operates through layers of abstraction that obscure the relationship between malicious action and harmful outcome. This complexity stems from the fundamental nature of machine learning systems, which aggregate data from countless sources, process information through opaque algorithms, and generate decisions that may only reveal their corrupted nature after deployment in real-world scenarios. The challenge extends beyond technical complexity to encompass questions of moral responsibility, legal culpability, and practical enforcement that existing legal frameworks struggle to address (Osmani, 2020).

The automated nature of AI decision-making creates what legal scholars have termed a “responsibility gap” in contemporary jurisprudence. When an AI system makes a harmful decision based on poisoned training data, identifying the responsible party requires tracing a causal chain through multiple stages of development, deployment, and operation. A healthcare AI that misdiagnoses patients due to corrupted training data exemplifies this challenge. The harm manifests through automated recommendations, but responsibility potentially rests with the attacker who poisoned the dataset, the developers who failed to detect the contamination, the organization that deployed the system without adequate testing, or the regulatory bodies that approved its use. Each actor in this chain may claim they acted reasonably given the information available, yet the aggregate result remains a serious public harm requiring legal accountability.

Proving criminal intent in AI poisoning cases confronts prosecutors with unprecedented difficulties rooted in the technical sophistication of these attacks. Traditional criminal law requires demonstrating that defendants possessed specific mental states when committing prohibited acts. A perpetrator must know their actions violate the law and intend the resulting harm. Yet AI poisoning can occur through data contributions that appear entirely legitimate on their surface. An attacker might submit seemingly normal images to a training dataset, each containing subtle pixel modifications imperceptible to human observers but capable of systematically corrupting the model's learned patterns. When prosecutors attempt to establish that such contributions were intentionally malicious rather than accidental errors or legitimate data variations, they face the burden of proving knowledge of highly technical attack methodologies and specific intent to cause harm through processes that may take months to manifest (Cheong et al., 2025).

The evidential challenges multiply when considering the distributed nature of modern AI development. Machine learning models typically train on datasets aggregated from numerous sources, with data collection, curation, and model training often performed by different entities. An autonomous vehicle manufacturer might purchase training data from multiple vendors, each collecting information from various sensors and annotators. If the resulting AI system exhibits dangerous behaviors traced to poisoned data, establishing which data source contained the corruption and whether its inclusion was intentional or negligent requires forensic capabilities that few law enforcement agencies possess. Prosecutors must not only identify the technical origin of contamination but also prove that responsible parties knew or should have known about the risk and failed to take adequate preventive measures.

The concept of negligence becomes particularly problematic in AI poisoning contexts because industry standards for data validation and model robustness remain poorly defined and rapidly evolving. Traditional negligence doctrine asks whether defendants exercised reasonable care according to prevailing professional standards. However, the AI field lacks consensus on what constitutes adequate testing for adversarial robustness or sufficient data validation to prevent poisoning attacks. A company might conduct extensive quality checks that satisfy current industry practice yet still deploy a poisoned model because detection techniques lag behind attack methodologies. Courts attempting to evaluate whether such companies acted negligently find themselves without clear benchmarks for reasonable care, forcing judges and juries to make highly technical judgments about emerging technologies without established legal guidance (Alnasser, 2025).

The multi-party nature of AI systems compounds liability questions by creating numerous potential defendants with varying degrees of responsibility. Consider a facial recognition system deployed by law enforcement that exhibits racial bias due to training

data poisoning. Potential liable parties include the attacker who corrupted the data, the data collection company that failed to detect contamination, the AI development firm that trained the model without adequate robustness testing, the vendor that marketed the system without disclosing its vulnerabilities, and the police department that deployed it without proper validation. Each party may bear partial responsibility, yet determining how to apportion legal liability requires courts to make novel judgments about the respective duties of data providers, model developers, system integrators, and end users in preventing AI harms. Existing product liability and negligence frameworks provide limited guidance because AI systems combine aspects of products, services, and professional expertise in ways that challenge traditional legal categories.

The technical opacity of machine learning models introduces additional complications for establishing causation between alleged misconduct and resulting harm. Modern deep learning systems function as black boxes where even their creators cannot fully explain how specific training examples influence particular predictions. When an AI system produces harmful outputs, proving that specific poisoned data caused those outputs rather than other factors requires sophisticated technical analysis that may be impossible with current forensic capabilities. Defense attorneys can exploit this uncertainty by arguing that harmful behaviors resulted from legitimate data variations, algorithmic limitations, or unforeseen interactions rather than deliberate poisoning. Prosecutors must overcome these arguments by presenting evidence that not only demonstrates correlation between suspected poisoned data and harmful outputs but establishes causation with the level of certainty required for criminal conviction (Wojtczak & Księżak, 2021).

The delayed manifestation of harm in AI poisoning cases creates statute of limitations problems that further complicate prosecution. Criminal statutes typically begin their limitations period when the offense occurs or when it is discovered. However, AI poisoning may involve data contributions made years before the trained model produces harmful outputs in deployment. An attacker might poison a dataset in early stages of model development, with the corrupted system only exhibiting dangerous behaviors after extensive additional training and deployment. Determining when the crime occurred and whether limitations periods have expired requires courts to decide whether the offense is the initial data corruption, the model training process, the system deployment, or the first harmful prediction. Different jurisdictions may adopt conflicting approaches, creating uncertainty that undermines effective prosecution.

The intersection of AI poisoning with international law raises additional liability complexities because development, deployment, and harm often occur in different jurisdictions with varying legal standards. A company headquartered in one country might train models using data collected globally, deploy the system through cloud infrastructure in another jurisdiction, and cause harm to users worldwide. If poisoning

occurs, prosecutors must navigate conflicting laws regarding data protection, algorithmic accountability, and criminal jurisdiction. Some countries impose strict liability for AI harms while others require proving fault. Some recognize AI poisoning as a distinct offense while others attempt to prosecute under general computer fraud statutes. These variations create opportunities for liable parties to exploit jurisdictional gaps by structuring operations to minimize legal exposure while making coordinated prosecution nearly impossible.

The corporate structure of AI development further obscures liability by distributing responsibility across subsidiaries, contractors, and third-party vendors. Large technology companies often develop AI systems through complex networks of entities, with parent corporations claiming limited liability for subsidiary actions while subsidiaries assert, they followed parent company guidance. When poisoning occurs, determining which corporate entity bears legal responsibility requires piercing through organizational structures designed to limit liability exposure. Prosecutors face the challenge of proving that specific individuals or entities within these networks knew about poisoning risks and failed to take adequate preventive action, all while company attorneys assert that responsibility diffused across the organization means no single party can be held liable (Novelli et al., 2024).

The rapid evolution of AI technology creates a moving target for liability standards as attack techniques and defensive capabilities evolve faster than legal precedents can develop. Courts deciding AI poisoning cases today must apply laws written before machine learning became widespread, relying on analogies to traditional cybercrimes that may not capture the unique characteristics of AI threats. By the time appellate courts establish liability principles for one type of poisoning attack, perpetrators may have developed new techniques that exploit different vulnerabilities. This temporal mismatch between legal development and technological change means that liability standards will perpetually lag behind current threats unless legislatures adopt more dynamic regulatory approaches.

The economic incentives surrounding AI liability create additional barriers to effective accountability. Companies developing AI systems face pressure to deploy quickly in competitive markets, creating temptations to minimize security testing that might delay product launches. The costs of comprehensive data validation and adversarial robustness testing can be substantial, particularly for startups and smaller companies with limited resources. If liability standards remain uncertain and enforcement weak, rational economic actors may conclude that the expected cost of potential legal consequences is lower than the guaranteed cost of thorough security measures. This dynamic creates a race to the bottom where companies minimize precautions, increasing overall vulnerability to poisoning attacks while making it harder to establish negligence when incidents occur.

The role of insurance in AI liability introduces further complications as insurers struggle to price risks for unprecedented threats with limited actuarial data. Companies increasingly purchase cyber insurance policies that may or may not cover AI poisoning incidents depending on policy language drafted before such attacks became prominent. When harm occurs, disputes over coverage can delay victim compensation while making it unclear whether insurance or company assets will satisfy liability claims. The uncertainty surrounding insurance coverage may also affect corporate incentives for security investments, with companies potentially relying on insurance protection rather than implementing robust safeguards (Aleksandrova et al., 2023).

The intersection of civil and criminal liability creates additional complexity in AI poisoning cases. While criminal prosecution requires proving guilt beyond reasonable doubt, civil liability typically applies lower standards of proof. Victims of AI poisoning might pursue civil damages even when criminal prosecution fails, creating parallel legal processes with potentially conflicting outcomes. A company might escape criminal charges yet face substantial civil liability, or vice versa. These divergent paths raise questions about whether civil liability can adequately deter AI poisoning or whether criminal penalties are necessary to address the severity of these threats. The answer likely depends on the specific context, with different liability mechanisms appropriate for different types of AI systems and harms.

The public policy implications of AI liability standards extend beyond individual cases to shape the broader trajectory of AI development. Overly strict liability could stifle innovation by making companies reluctant to develop beneficial AI applications due to fear of legal exposure. Conversely, insufficient liability allows harmful systems to proliferate without adequate accountability. Finding the appropriate balance requires policymakers to consider not only legal principles but also technological feasibility, economic impacts, and societal values. The challenge is particularly acute given the global nature of AI development, which means that liability standards adopted in one jurisdiction affect competitive dynamics worldwide.

Recent legislative efforts attempt to address these liability challenges through new regulatory frameworks specifically designed for AI systems. The European Union's AI Act establishes a risk-based approach where high-risk AI systems face strict requirements for data governance, documentation, and testing, with penalties for non-compliance. However, these regulations focus primarily on organizational compliance rather than individual criminal liability for poisoning attacks. The United States has taken a more fragmented approach with sector-specific regulations and state-level initiatives rather than comprehensive federal legislation. These divergent regulatory strategies reflect deeper disagreements about how to balance innovation incentives with accountability imperatives.

The development of technical standards for AI security represents a crucial

complement to legal liability frameworks. Organizations like the National Institute of Standards and Technology have begun developing guidelines for adversarial robustness testing and data validation that could inform legal standards of care. If courts adopt these technical standards as benchmarks for negligence determinations, they provide clearer guidance about required precautions while allowing standards to evolve with technological capabilities. However, this approach depends on technical standards keeping pace with emerging threats and gaining broad industry acceptance, neither of which is guaranteed (Mansouri et al., 2025).

The role of expert testimony in AI poisoning cases introduces another layer of complexity as judges and juries must evaluate highly technical evidence about machine learning vulnerabilities, attack methodologies, and defensive capabilities. The adversarial nature of litigation means that competing experts may present contradictory interpretations of the same technical evidence, leaving factfinders without clear guidance. Courts have struggled with junk science in other technical domains, and AI poisoning cases risk similar problems if expert standards are not carefully developed. Ensuring that liability determinations rest on sound technical foundations requires investing in judicial education and developing clear standards for admissible AI expert testimony.

### C. Jurisdictional Barriers in Prosecuting Cross-Border AI Model Poisoning

The global nature of AI development and deployment creates profound challenges for law enforcement agencies attempting to investigate and prosecute model poisoning attacks. Unlike traditional crimes that occur within defined geographic boundaries, AI poisoning exploits the inherently distributed architecture of modern machine learning systems. Training data originates from multiple countries, computational infrastructure spans international cloud networks, development teams work across continents, and deployed systems affect users worldwide. This distributed reality clashes fundamentally with territorial principles that underpin most criminal justice systems, where jurisdiction depends on where crimes occur or where defendants reside. The resulting enforcement gaps allow perpetrators to operate with relative impunity by strategically positioning their activities across jurisdictional boundaries that investigators struggle to cross (Wisnubroto & Hilaire Tegnan, 2025).

International legal cooperation mechanisms developed for traditional cybercrimes prove inadequate when applied to AI poisoning cases. The Budapest Convention on Cybercrime, adopted in 2001, represents the primary international framework for cross-border cooperation on computer-related offenses. Over sixty countries have ratified this treaty, establishing protocols for mutual legal assistance, extradition, and coordinated investigations. However, the Convention addresses traditional hacking activities like unauthorized system access and data theft rather than sophisticated manipulation of machine learning algorithms. AI poisoning attacks often involve no illegal access to systems but instead corrupt publicly available datasets or exploit legitimate data

contribution channels. Prosecutors attempting to invoke Convention procedures for AI poisoning cases find that treaty language fails to clearly encompass these activities, creating uncertainty about whether requesting states can compel cooperation from treaty partners.

The technical complexity of AI poisoning investigations exceeds the capabilities of mutual legal assistance procedures designed for simpler cybercrimes. When law enforcement in one country suspects that training data corruption originated from another jurisdiction, requesting assistance requires explaining highly technical attack methodologies to foreign authorities who may lack specialized AI expertise. The requesting state must articulate what evidence it seeks, but identifying relevant evidence in AI poisoning cases demands understanding machine learning architectures, data provenance tracking, and adversarial attack techniques. Foreign authorities receiving assistance requests may struggle to comprehend what investigators need or how to obtain it from local technology companies (Stoykova et al., 2024). These communication barriers delay investigations, allowing perpetrators to destroy evidence or continue their attacks while authorities negotiate technical details across language and expertise divides.

Jurisdictional conflicts arise when multiple countries claim authority to prosecute the same AI poisoning incident, creating risks of double jeopardy and conflicting legal outcomes. Consider an attack where perpetrators in Country A poison training data hosted on servers in Country B, corrupting a model developed by a company headquartered in Country C that deploys the system in Country D, ultimately causing harm to users in Country E. Each jurisdiction may assert legitimate grounds for prosecution based on different connecting factors. The principle of territoriality suggests Countries B, C, and D have jurisdiction based on where criminal acts or their effects occurred. The nationality principle allows Country A to prosecute its citizens regardless of where crimes occurred. The passive personality principle permits Country E to prosecute based on victim nationality. Without clear international protocols for resolving these competing claims, investigations may proceed in parallel with inefficient duplication of effort or may stall entirely as countries dispute who should take the lead.

The absence of harmonized legal definitions for AI-related offenses compounds cross-border enforcement difficulties. Countries that criminalize AI poisoning use vastly different statutory language and required elements of proof. Some jurisdictions classify poisoning as computer fraud, others as sabotage, and still others create specific AI manipulation offenses. These definitional variations mean that conduct constituting a serious crime in one country may not be criminal at all in another. When investigators seek assistance from countries that do not criminalize the conduct under investigation, requests may be denied based on the principle of dual criminality, which requires that alleged conduct be criminal in both requesting and requested states. Perpetrators exploit these definitional gaps by locating their operations in jurisdictions that do not recognize

AI poisoning as criminal, effectively creating safe havens from which to launch attacks with minimal legal risk.

Extradition challenges further undermine cross-border enforcement efforts when suspects refuse to voluntarily appear in jurisdictions seeking to prosecute them. Extradition treaties typically require showing probable cause that the suspect committed an extraditable offense and that the offense is criminal in both countries. For novel AI poisoning cases, establishing probable cause demands presenting technical evidence that foreign judges may find difficult to evaluate. Defense attorneys challenge extradition requests by arguing that AI poisoning does not fall within treaty definitions of extraditable offenses or that allegations are politically motivated persecution of legitimate AI research. These arguments find receptive audiences in countries protective of their technology sectors or reluctant to extradite citizens for offenses poorly defined in international law. The result is that identified suspects often remain beyond the reach of jurisdictions most affected by their attacks (Button et al., 2025).

Data protection regulations create additional cross-border enforcement obstacles by restricting the flow of information that investigators need to trace AI poisoning attacks. The European Union's General Data Protection Regulation imposes strict limitations on transferring personal data outside the EU, even for law enforcement purposes. When investigators in non-EU countries need access to training data, model parameters, or user information to trace the origins and impacts of poisoning attacks, GDPR restrictions may prevent EU-based companies from providing this information without complex legal procedures. Similar data localization requirements in countries like China, Russia, and India create a fragmented global landscape where evidence necessary for comprehensive investigations remains trapped within jurisdictional silos. Companies operating internationally must navigate conflicting obligations, sometimes facing the impossible choice between violating data protection laws or obstructing criminal investigations.

The commercial sensitivity of AI systems adds another dimension to cross-border enforcement challenges as companies resist disclosing proprietary information even when legally compelled. Machine learning models represent substantial investments in research and development, with their architectures and training data constituting valuable trade secrets. When investigators request access to examine systems for evidence of poisoning, companies worry that disclosure risks competitive harm if information leaks or must be shared in court proceedings. These concerns intensify in cross-border contexts where legal protections for confidential information vary significantly. A company may trust domestic courts to safeguard trade secrets but fear that foreign legal systems lack adequate protections. Consequently, companies employ litigation strategies to delay or limit disclosure, impeding investigations that depend on timely access to technical evidence (Geraldine O Mbah, 2024).

Resource constraints affecting law enforcement agencies create severe practical barriers to effective cross-border AI poisoning investigations. Pursuing international cases demands specialized expertise in machine learning, foreign language capabilities, diplomatic coordination skills, and sustained funding for travel and technical analysis. Most law enforcement agencies lack dedicated units with these combined competencies. Investigators trained in traditional cybercrime may understand network forensics but struggle with adversarial machine learning concepts. AI specialists may lack law enforcement experience and legal knowledge. Building international coalitions for specific investigations requires investing significant time in relationship building and procedural coordination. Given competing demands on limited resources, agencies often prioritize domestic cases with clearer legal frameworks over complex international AI poisoning investigations with uncertain prospects for successful prosecution.

Recent initiatives attempt to strengthen international cooperation on AI-related crimes, though their effectiveness remains uncertain. INTERPOL has established a cybercrime directorate that includes AI security within its mandate, providing a platform for information sharing among member countries. The United Nations has convened expert groups to discuss international legal frameworks for emerging technologies, including AI security threats. Regional organizations like the European Union and African Union have begun developing coordinated approaches to AI governance that include criminal enforcement dimensions. However, these initiatives face the same fundamental challenge of reconciling diverse legal systems, competing national interests, and rapid technological change. Progress occurs slowly through consensus-building processes while AI capabilities and associated threats evolve at accelerating rates.

The geopolitical dimensions of AI competition further complicate international enforcement cooperation. Countries view AI capabilities as strategic assets essential for economic competitiveness and national security. This perspective creates incentives to protect domestic AI industries from foreign legal actions that might disadvantage national champions. When poisoning allegations involve researchers or companies from competing nations, governments may suspect political motivations behind enforcement actions. The United States, China, and European Union pursue divergent AI governance philosophies reflecting different values regarding privacy, security, and innovation. These divergent approaches undermine the shared normative foundation necessary for effective international legal cooperation. Building trust across geopolitical divides requires diplomatic efforts extending far beyond technical legal cooperation mechanisms.

### D. Legal Unpreparedness for Prosecuting AI Model Poisoning as Cybercrime

The scarcity of judicial decisions addressing AI model poisoning creates profound uncertainty throughout legal systems attempting to respond to these emerging threats. Courts rely heavily on precedent to interpret statutes, establish procedural standards, and develop legal doctrines that provide predictable frameworks for future cases. When novel

situations arise without relevant precedents, judges must reason by analogy from existing case law, often applying legal principles developed for fundamentally different contexts. AI model poisoning presents precisely this challenge, involving technical complexities and causal relationships that bear little resemblance to the traditional cybercrimes that dominate existing jurisprudence. The resulting legal vacuum leaves prosecutors uncertain about what charges to bring, defense attorneys without established arguments to counter novel theories, and judges lacking guidance on how to evaluate evidence or instruct juries about highly technical concepts (Nastoska et al., 2025).

The few reported cases involving AI security issues rarely address model poisoning directly but instead focus on related concerns like algorithmic bias, data breaches, or intellectual property theft. When courts encounter AI systems that produce discriminatory outputs or make erroneous decisions, they typically analyze these issues through existing frameworks for employment discrimination, consumer protection, or negligence rather than recognizing poisoning as a distinct criminal offense. This pattern reflects both the difficulty of proving intentional poisoning versus other causes of AI failures and the reluctance of prosecutors to pursue novel legal theories when traditional charges might succeed. However, treating all AI malfunctions through conventional legal lenses prevents the development of jurisprudence specifically addressing the unique characteristics of adversarial attacks on machine learning systems.

The technical opacity of AI poisoning attacks contributes significantly to the absence of case law by making these incidents difficult to detect, investigate, and prove in court. Unlike traditional cybercrimes that leave clear digital footprints through unauthorized access logs or system modifications, poisoning attacks can appear indistinguishable from legitimate data contributions. An attacker who subtly corrupts training data by submitting carefully crafted examples may leave no evidence of malicious intent that investigators can discover. Even when suspicious patterns emerge, linking them definitively to specific actors requires sophisticated forensic analysis that exceeds the capabilities of most law enforcement agencies. Consequently, many poisoning incidents likely go undetected or remain unresolved, never reaching courts to generate precedential decisions that could guide future cases (Radanliev, 2025).

When AI poisoning issues do reach litigation, they typically arise in civil contexts rather than criminal prosecutions, limiting the development of criminal law precedents. Companies harmed by compromised AI systems may sue vendors, service providers, or contractors for breach of contract, negligence, or fraud. These civil cases focus on compensating victims rather than punishing perpetrators, applying contractual interpretation and tort principles rather than criminal statutes. While civil precedents may inform criminal prosecutions by establishing factual findings about how poisoning occurred or what harms resulted, they provide limited guidance on criminal elements like

intent, jurisdiction, or sentencing. The predominance of civil over criminal cases reflects both the difficulties of meeting criminal burden of proof standards and the private sector's preference for resolving disputes through commercial litigation rather than involving law enforcement.

The rapid evolution of AI technology creates a moving target that prevents precedents from accumulating into coherent legal frameworks. By the time courts decide cases involving particular attack methodologies or system architectures, technology has advanced, rendering the specific circumstances addressed in those decisions less relevant to subsequent cases. A precedent involving poisoning of a simple image classifier may offer limited guidance for cases involving large language models or reinforcement learning systems with fundamentally different vulnerabilities and operational characteristics. This temporal mismatch between judicial and technological timescales means that case law perpetually lags behind current threats, forcing each new generation of cases to confront novel questions without adequate precedential foundations.

Judicial unfamiliarity with machine learning concepts compounds the challenges of establishing useful precedents even when cases do reach courts. Judges trained in traditional legal analysis may struggle to evaluate competing expert testimony about adversarial robustness, gradient-based attacks, or data provenance verification. Without deep technical understanding, courts may mischaracterize key facts or adopt reasoning that reflects fundamental misconceptions about how AI systems function. These errors become embedded in precedent, potentially misleading future courts and creating doctrine built on flawed technical foundations. The problem intensifies at appellate levels where judges even further removed from technical details must review trial court decisions and establish broader legal principles. Ensuring that emerging AI case law rests on sound technical understanding requires significant investments in judicial education and improved mechanisms for conveying complex technical information in legal proceedings (Qutieshat et al., 2024).

The settlement of cases before trial prevents many potential precedent-setting disputes from generating published judicial opinions. Companies involved in AI poisoning incidents face strong incentives to resolve matters quietly through confidential settlements rather than risk public trials that might reveal security vulnerabilities, damage reputations, or establish unfavorable precedents. Prosecutors considering criminal charges may accept plea agreements rather than proceeding to trial when defendants offer cooperation or when uncertainties about novel legal theories create risks of acquittal. While settlements and plea agreements efficiently resolve individual cases, they deprive the legal system of opportunities to develop publicly available precedents that could guide future disputes. The cumulative effect is a body of hidden case outcomes that might inform legal understanding if accessible but instead remains locked away in confidential

agreements.

International variations in legal systems further fragment the already limited case law on AI poisoning across jurisdictions with incompatible precedential structures. Common law systems like those in the United States, United Kingdom, and former British colonies rely heavily on judicial precedent as a primary source of law. Civil law jurisdictions throughout Europe, Latin America, and Asia give precedent less formal weight, with courts focusing more on statutory interpretation and legal scholarship. These structural differences mean that even when courts in various countries address similar AI poisoning issues, the resulting decisions accumulate into separate bodies of national case law rather than converging toward international consensus. A precedent established in American courts may influence other common law jurisdictions but carries little weight in civil law countries and vice versa. This fragmentation prevents the emergence of globally consistent legal approaches to inherently transnational threats.

The absence of precedent creates practical difficulties for all participants in potential AI poisoning litigation. Prosecutors cannot confidently assess which legal theories courts will accept, making charging decisions risky when novel approaches might fail completely. Defense attorneys lack established arguments for challenging prosecution theories or defending clients accused under untested statutes. Judges must make critical decisions about admissibility of evidence, jury instructions, and sentencing without guidance from prior cases addressing comparable situations. Juries receive little help understanding technical concepts when judges themselves struggle to explain machine learning principles in comprehensible terms. These uncertainties disadvantage all parties and risk producing arbitrary outcomes that depend more on particular judges' intuitions than on consistent application of legal principles (Moch, 2024).

The lack of precedent also undermines deterrence by leaving potential perpetrators uncertain about legal consequences of AI poisoning activities. Criminal law achieves deterrent effects partly through clearly communicating what conduct is prohibited and what punishments will follow. When case law remains undeveloped, individuals considering poisoning attacks cannot reliably assess their legal risks. Some may refrain from clearly criminal conduct out of caution, but others may proceed with attacks believing that legal uncertainties make prosecution unlikely or that novel defenses might succeed. Similarly, companies deciding how much to invest in security measures lack clear signals about what precautions courts will deem adequate to avoid liability. This uncertainty potentially leads to both excessive caution that stifles beneficial innovation and inadequate precautions that leave systems vulnerable.

Recent efforts to address the precedent gap include specialized training programs for judges, the creation of technology courts in some jurisdictions, and increased use of technical advisors in complex cases. Several countries have established dedicated

intellectual property or technology courts where judges develop expertise in technical subjects through repeated exposure to similar cases. The United States federal court system has experimented with appointing technical experts to assist judges in understanding complex evidence, though questions remain about how to select neutral experts and what roles they should play in judicial decision-making. Legal scholars have begun systematically analyzing the limited AI case law that exists, attempting to extract principles that might guide future courts even when precedents address somewhat different technical contexts.

Academic commentary and legal scholarship play particularly important roles in shaping AI law during this precedent-deficit period. Law review articles, treatises, and practice guides attempt to fill gaps left by absent case law by proposing frameworks for analyzing AI poisoning cases, predicting how courts might resolve novel questions, and recommending legislative reforms. Courts sometimes cite academic sources when deciding cases of first impression, giving scholarship unusual influence in emerging legal fields. However, academic analysis cannot fully substitute for judicial precedent because scholarly proposals lack the binding authority of court decisions and may reflect theoretical perspectives divorced from practical realities of litigation. Nevertheless, quality legal scholarship helps frame issues, identify relevant analogies, and develop vocabulary for discussing AI security threats in legal contexts.

### E. Implication

This research fundamentally questions traditional cybercrime theories that view digital attacks as discrete events involving unauthorized system access. AI model poisoning operates through gradual corruption during legitimate interactions, requiring new conceptual frameworks recognizing harm from accumulated data manipulation rather than single intrusive acts. Existing deterrence theories fail when perpetrators exploit open data channels without violating access controls. Classical attribution models assuming traceable footprints prove inadequate when malicious actors submit poisoned data anonymously.

The findings create pathways for enhanced international cooperation by identifying treaty gaps preventing effective responses. Policymakers gain insights for designing legislation addressing novel threats while preserving innovation incentives. Technology companies benefit from clearer legal expectations, enabling strategic security investments. However, heightened awareness may encourage attacks in jurisdictions lacking specific statutes. Overly aggressive responses risk chilling legitimate research, particularly adversarial machine learning studies improving system robustness. Small startups face disproportionate compliance burdens compared to large corporations.

Lawmakers should craft statutes explicitly defining AI poisoning with clear

elements distinguishing malicious attacks from negligent data issues. The European Union's AI Act provides a risk-based model for high-stakes systems. Organizations must establish data provenance tracking documenting training dataset origins and integrity. Security teams need continuous monitoring for anomalous behaviors suggesting poisoning beyond traditional network security. Law enforcement requires specialized units combining machine learning expertise with investigative skills. These comprehensive reforms balance accountability with innovation, ensuring AI systems serve public interests while maintaining security against emerging threats.

### F. Recommendations

Governments should establish specialized AI crime units within national cybersecurity agencies equipped with machine learning forensic capabilities and dedicated prosecution teams. International bodies must draft a comprehensive AI Security Convention explicitly addressing model poisoning, data corruption, and algorithmic manipulation across borders. Countries should implement mandatory reporting requirements for AI poisoning incidents affecting critical infrastructure, similar to data breach notification laws already operating in healthcare and finance sectors. Technology companies need industry-wide standards for adversarial robustness testing before deploying AI systems in high-risk domains. Universities and training academies should develop certification programs for judges, prosecutors, and investigators focused on AI-specific evidentiary challenges and technical concepts.

Existing cybercrime frameworks require amendment to recognize data integrity violations as distinct offenses separate from unauthorized access crimes. The Budapest Convention needs updating to explicitly include AI poisoning within its scope and establish streamlined mutual assistance protocols. Current liability models treating AI developers, deployers, and data providers as separate entities should shift toward shared responsibility frameworks reflecting collaborative development processes. Evidence rules must accommodate novel forensic techniques for tracing poisoned data through complex training pipelines. Sentencing guidelines should account for the delayed and widespread harms characteristic of AI attacks rather than applying penalties designed for immediate, localized cybercrimes.

This research acknowledges constraints including limited access to confidential corporate incident data and rapidly evolving attack methodologies that may outpace analysis. Future investigations should examine sector-specific vulnerabilities in healthcare, autonomous vehicles, and financial AI systems. Researchers must explore technical solutions like cryptographic data verification and blockchain-based provenance tracking. Comparative studies analyzing regulatory approaches across jurisdictions will identify best practices. Empirical research quantifying the deterrent effects of various

legal frameworks remains critically needed.

## Conclusion

AI model poisoning represents an urgent threat that existing legal systems struggle to address effectively. Traditional cybercrime laws were designed for direct attacks on computer networks, not subtle manipulation of machine learning algorithms through corrupted training data. As artificial intelligence becomes embedded in healthcare diagnostics, autonomous vehicles, financial systems, and criminal justice, the capacity to secretly corrupt these systems poses risks to public safety and social trust. Current regulatory frameworks lack clear definitions of AI-specific offenses, leaving prosecutors uncertain about applicable charges and perpetrators operating with minimal fear of consequences. The global nature of AI development amplifies these challenges as attacks span multiple jurisdictions while international cooperation mechanisms remain inadequate for coordinating investigations across borders.

The research reveals interconnected obstacles preventing effective accountability. Legal systems cannot establish liability when automated processes obscure causal chains between data corruption and harmful outcomes. Proving criminal intent becomes nearly impossible when attacks involve seemingly legitimate data contributions containing imperceptible modifications. Cross-border enforcement fails because existing treaties like the Budapest Convention do not explicitly address AI poisoning, creating jurisdictional gaps that attackers exploit strategically. Courts lack precedents for interpreting novel statutes or evaluating technical evidence about adversarial machine learning, resulting in inconsistent outcomes that undermine legal predictability. These findings collectively demonstrate that incremental reforms cannot suffice; comprehensive legislative action specifically targeting AI security threats is essential.

Cybercrimes issues and challenges requires coordinated action across multiple domains. Governments must enact statutes explicitly criminalizing AI poisoning with clear elements distinguishing malicious attacks from negligent errors. International bodies should draft treaties establishing streamlined cooperation protocols for AI crime investigations. Technology companies need mandatory security standards ensuring robust testing before deploying systems in critical applications. Law enforcement agencies require specialized training in machine learning forensics. Future research should examine sector-specific vulnerabilities, evaluate deterrent effects of various legal approaches, and explore technical solutions like cryptographic data verification that complement regulatory frameworks.

## Bibliography

Aleksandrova, A., Ninova, V., & Zhelev, Z. (2023). A Survey on AI Implementation in Finance, (Cyber) Insurance and Financial Controlling. *Risks*, *11*(5), 91. https://doi.org/10.3390/risks11050091

Allheeib, N. (2024). Securing Machine Learning Against Data Poisoning Attacks. *International Journal of Data Warehousing and Mining*, *20*(1), 1–21. https://doi.org/10.4018/IJDWM.358335

Alnasser, H. A. (2025). THE CONCEPT OF NEGLIGENCE IN DATA BREACH: A COMPARATIVE DOCTRINAL ANALYSIS OF THE EU, CALIFORNIA, AND SAUDI ARABIA. *Veredas Do Direito*, *22*(3), e223404. https://doi.org/10.18623/rvd.v22.n3.3404

Button, M., Hock, B., Suh, J. B., & Koh, C. S. (2025). Policing cross-border fraud 'Above and below the surface': mapping actions and developing a more effective global response. *Crime, Law and Social Change*, *83*(1), 5. https://doi.org/10.1007/s10611-024-10186-2

Cheong, I., Caliskan, A., & Kohno, T. (2025). Safeguarding human values: rethinking US law for generative AI's societal impacts. *AI and Ethics*, *5*(2), 1433–1459. https://doi.org/10.1007/s43681-024-00451-4

Cotroneo, D., Improta, C., Liguori, P., & Natella, R. (2024). Vulnerabilities in AI Code Generators: Exploring Targeted Data Poisoning Attacks. *Proceedings of the 32nd IEEE/ACM International Conference on Program Comprehension*, 280–292. https://doi.org/10.1145/3643916.3644416

Diro, A., Kaisar, S., Saini, A., Fatima, S., Hiep, P. C., & Erba, F. (2025). Workplace security and privacy implications in the GenAI age: A survey. *Journal of Information Security and Applications*, *89*, 103960. https://doi.org/10.1016/j.jisa.2024.103960

Geraldine O Mbah. (2024). Data privacy in the era of AI: Navigating regulatory landscapes for global businesses. *International Journal of Science and Research Archive*, *13*(2), 2040–2058. https://doi.org/10.30574/ijsra.2024.13.2.2396

Iyer, K. I. (2023). Poisoning AI Models: New Frontiers in Data Manipulation Attacks. *International Journal of Innovative Research in Computer and Communication Engineering*, *11*(11). https://doi.org/10.15680/IJIRCCE.2023.1111065

Mansouri, O., Yusuf, N., & Kooli, C. (2025). Ethical frontiers and legal boundaries: Proposing a unified framework for AI regulation and accountability. *Next Research*, *2*(4), 101087. https://doi.org/10.1016/j.nexres.2025.101087

Moch, E. (2024). Liability Issues in the Context of Artificial Intelligence: Legal Challenges and Solutions for AI-Supported Decisions. *East African Journal of Law and Ethics*, *7*(1), 214–234. https://doi.org/10.37284/eajle.7.1.2518

Nastoska, A., Jancheska, B., Rizinski, M., & Trajanov, D. (2025). Evaluating Trustworthiness in AI: Risks, Metrics, and Applications Across Industries. *Electronics*, *14*(13), 2717. https://doi.org/10.3390/electronics14132717

Novelli, C., Casolari, F., Hacker, P., Spedicato, G., & Floridi, L. (2024). Generative AI in EU law: Liability, privacy, intellectual property, and cybersecurity. *Computer Law & Security Review*, *55*, 106066. https://doi.org/10.1016/j.clsr.2024.106066

Osmani, N. (2020). The Complexity of Criminal Liability of AI Systems. *Masaryk University Journal of*

*Law and Technology*, *14*(1), 53–82. https://doi.org/10.5817/MUJLT2020-1-3

Panattoni, B. (2025). Generative AI and Criminal Guilt. In *The Cambridge Handbook of Generative AI and the Law* (pp. 392–404). Cambridge University Press. https://doi.org/10.1017/9781009492553.027

Qutieshat, E. M. A., Quteishat, A. M. A., & Qtaishat, A. (2024). Transforming the Judicial System: The Impact of Machine Learning on Legal Processes and Outcomes. *International Journal of Religion*, *5*(11), 6833–6841. https://doi.org/10.61707/9gtxnr11

Radanliev, P. (2025). Frontier AI regulation: what form should it take? *Frontiers in Political Science*, *7*. https://doi.org/10.3389/fpos.2025.1561776

Sarkar, G., & Shukla, S. K. (2023). Behavioral analysis of cybercrime: Paving the way for effective policing strategies. *Journal of Economic Criminology*, *2*, 100034. https://doi.org/10.1016/j.jeconc.2023.100034

Stoykova, R., Porter, K., & Beka, T. (2024). The AI Act in a law enforcement context: The case of automatic speech recognition for transcribing investigative interviews. *Forensic Science International: Synergy*, *9*, 100563. https://doi.org/10.1016/j.fsisyn.2024.100563

Sun, J., Gu, S., & Su, R. (2026). AI-Empowered Responsive Regulation for Preventing Future Crimes: An Empirical Inquiry into the Regulatory Pyramid to Combat Future Crimes in China and Southeast Asia. *Asian Journal of Criminology*, *21*(1), 8. https://doi.org/10.1007/s11417-025-09477-x

Wisnubroto, A., & Hilaire Tegnan. (2025). Preventing AI Crime Towards A New Legal Paradigm: Lessons From United States. *Journal of Human Rights, Culture and Legal System*, *5*(2), 630–658. https://doi.org/10.53955/jhcls.v5i2.606

Wojtczak, S., & Księżak, P. (2021). Causation in Civil Law and the Problems of Transparency in AI. *European Review of Private Law*, *29*(Issue 4), 561–582. https://doi.org/10.54648/ERPL2021030

Zaidan, E., & Ibrahim, I. A. (2024). AI Governance in a Complex and Rapidly Changing Regulatory Landscape: A Global Perspective. *Humanities and Social Sciences Communications*, *11*(1), 1121. https://doi.org/10.1057/s41599-024-03560-x

Zhang, R., Li, H.-W., Qian, X.-Y., Jiang, W.-B., & Chen, H.-X. (2025). On large language models safety, security, and privacy: A survey. *Journal of Electronic Science and Technology*, *23*(1), 100301. https://doi.org/10.1016/j.jnlest.2025.100301