

The Criminal Law Challenges in Confronting AI Crimes تحديات القانون الجنائي في مواجهة الجرائم المرتبطة بالذكاء الاصطناعي



Tayeb Belouadah^{1*}

¹ Faculty of Law and Political Sciences, University of M'sila, Algeria

tayeb.belouadah@univ-msila.dz

Date of transmission: 07/05/2025 date of acceptance: 29/05/2025 date of publication: 01/06/2025

Abstract:

In this research, we address the challenges posed by artificial intelligence to the criminal law system, where we classify the various forms of AI-supported criminal activities and identify the associated judicial and territorial obstacles. We then analyze the criminal responsibility gap that arises when autonomous AI systems commit harmful acts without clear human intent or negligence. We also review the use of AI-based risk assessment tools within the criminal system itself, highlighting concerns regarding reliability, bias, and the balance between human and algorithmic decision-making. Additionally, we highlight the legal frameworks dedicated to dealing with AI crimes, and finally, we offer proposals for legal reforms that balance technological innovation on one hand with accountability and the protection of fundamental rights on the other.

key words: Criminal Law; Challenges; Confronting; AI Crimes.

المخلص

نتناول في هذا البحث التحديات التي يفرضها الذكاء الاصطناعي على منظومة القانون الجنائي، حيث نقوم بتصنيف الأشكال المختلفة للأنشطة الإجرامية المدعومة بالذكاء الاصطناعي، وتحديد العقوبات القضائية والإقليمية المرتبطة بها. ثم نحلل فجوة المسؤولية الجنائية التي تنشأ عندما ترتكب أنظمة ذكاء اصطناعي مستقلة أفعالاً ضارة دون وجود نية أو إهمال بشري واضح. كما نستعرض استخدام أدوات تقييم المخاطر المدعومة بالذكاء الاصطناعي داخل النظام الجنائي ذاته، مع تسليط الضوء على مخاوف تتعلق بالمصداقية والتحيز والتوازن بين القرار البشري والآلي. ونبرز الأطر القانونية المخصصة للتعامل مع جرائم الذكاء الاصطناعي، ونقدم في الأخير اقتراحات لإصلاحات قانونية تُوازن بين الابتكار التكنولوجي من جهة وبين المساءلة وحماية الحقوق الأساسية من جهة أخرى.

الكلمات المفتاحية: القانون الجنائي، تحديات، مواجهة الجرائم، الذكاء الاصطناعي

* Corresponding Author

Introduction:

In recent years, artificial intelligence (AI) has rapidly evolved from a theoretical concept to a ubiquitous technology embedded in countless aspects of modern society. This technological revolution has brought unprecedented benefits across various sectors, from healthcare to transportation, finance to entertainment. However, as AI systems become more sophisticated and autonomous, they also present novel challenges for legal frameworks designed in an era before such technologies were conceivable. The criminal justice system, in particular, faces significant hurdles in addressing what can be termed “AI crimes”—criminal activities either facilitated by AI or, more problematically, committed by autonomous AI systems themselves.

The intersection of AI and criminal law represents a complex frontier where traditional legal doctrines encounter technological realities that strain their foundational principles. Criminal law has historically been predicated on concepts of human agency, intent, and moral culpability. Yet, as AI systems increasingly operate with degrees of autonomy and opacity that obscure direct human control, these fundamental tenets face unprecedented challenges. When an AI trading algorithm autonomously manipulates financial markets, when a deepfake video creates false evidence, or when an autonomous vehicle causes a fatal accident, the question of who—or what—bears criminal responsibility becomes dauntingly complex.

These challenges are not merely theoretical concerns for future consideration but pressing issues demanding immediate attention. As noted by Nanos, “Future criminal law must be prepared and probably adjusted to effectively tackle any questions concerning criminal liability of artificial intelligence.”¹ The rapid pace of AI development has outstripped the evolution of legal frameworks designed to govern its use and misuse. This gap creates a precarious situation where harmful AI-enabled activities may fall outside the reach of existing criminal statutes, potentially leaving victims without recourse and society without adequate protections.

The criminal justice system’s response to AI crimes is further complicated by the technical complexity of these technologies. As Claypoole observes, “Different functions of artificial intelligence create different problems for law and society.”² From generative AI creating convincing deepfakes to decisioning AI making potentially biased determinations in criminal risk assessment, each application presents distinct legal challenges that resist one-size-fits-all

¹Andreas Nanos. “Criminal Liability of Artificial Intelligence.” Charles University in Prague Faculty of Law Research Paper (No. 2023/III/3, 4 Nov. 2023), p. 1.

²Claypoole, Theodore Franklin. “AI Classifications for Law and Regulation.” American Bar Association, 15 Sept. 2023, p. 3.

solutions. Judges, prosecutors, defense attorneys, and legislators often lack the technical expertise to fully comprehend the capabilities and limitations of AI systems implicated in criminal proceedings.

This law review article examines the multifaceted challenges that artificial intelligence poses to criminal law frameworks and explores potential pathways for addressing these challenges. It begins by categorizing the various forms of AI-enabled criminal activities and identifying the jurisdictional and territorial hurdles they present. The article then analyzes the “criminal responsibility gap” that emerges when autonomous AI systems engage in harmful behaviors without clear human intent or negligence. Next, it examines the use of AI risk assessment tools within the criminal justice system itself, highlighting concerns regarding validity, bias, and the appropriate balance between human and algorithmic decision-making. The article further explores emerging legal frameworks specifically designed to address AI-specific crimes such as deepfakes and algorithmic discrimination. Finally, it concludes with recommendations for legal and policy reforms that balance technological innovation with accountability and protection of fundamental rights.

As society continues to integrate increasingly autonomous AI systems into critical functions, the criminal justice system must evolve to address the novel challenges these technologies present. This evolution requires not only technical understanding but also careful consideration of the ethical, social, and legal implications of attributing criminal liability in an age of artificial intelligence. The stakes are significant: failure to adapt criminal law to the realities of AI could undermine public trust in both the justice system and beneficial AI technologies, while overly restrictive approaches could stifle innovation and its attendant benefits. Finding the appropriate balance demands thoughtful analysis and interdisciplinary collaboration—a process this article aims to advance.

Section I: Categorizing AI Crimes and Legal Challenges

In this section we discuss Typology of AI-Enabled Criminal Activities and Jurisdictional and Territorial Challenges.

A) Typology of AI-Enabled Criminal Activities

The landscape of AI-enabled criminal activities is diverse and rapidly evolving, presenting significant challenges for criminal justice systems worldwide. As noted by TRM Labs, AI crimes can be classified into three distinct phases of maturity: “horizon, emerging, and mature—illustrating how AI is amplifying criminal capabilities by removing traditional human

bottlenecks.”¹ This classification framework provides valuable insight into the evolving nature of AI crimes and their potential impact on society.

At the most basic level, AI can serve as a tool for enhancing traditional criminal activities. Cybercriminals are increasingly leveraging AI to automate attacks, with “AI-powered tools enabling the automation of phishing campaigns, creating highly convincing messages at scale.”² These tools allow criminals to operate more efficiently and at greater scale than previously possible, while simultaneously making their activities harder to detect and attribute. The U.S. Treasury Department has specifically highlighted how “existing threat actors use Generative AI to develop and pilot more sophisticated malware—giving them complex attack capabilities previously available only to the most well-resourced actors.”³

Beyond serving as a tool for traditional crimes, AI has enabled entirely new categories of criminal activity. Deepfakes represent one of the most concerning examples, with criminals employing “deepfake technology to impersonate executives or public figures, and AI-generated images and voices that make scams harder to detect.”⁴ These technologies facilitate high-value fraud, extortion, and social manipulation in ways that were technically impossible before recent advances in generative AI. The legal response to these threats has been fragmented, with “a minority of states adopting laws specifically addressing the nonconsensual dissemination of sexually explicit deepfakes.”⁵

Perhaps most concerning are scenarios involving autonomous AI criminal behavior, what Nerantzi and Sartor term “Hard AI Crime.” They define this as occurring “whenever the harmful consequence brought about by such an AI agent—an agent acting as a utility maximiser—was not intended by any of the humans who contributed to the agent’s design and deployment.”⁶ In such cases, an AI system might autonomously engage in behavior that constitutes a crime, such as market manipulation, without any human having specifically intended or foreseen this outcome. This creates what the authors describe as a “criminal

¹ TRM Labs. “The Rise of AI-Enabled Crime: Exploring the evolution, risks, and responses to AI-powered criminal enterprises.” (TRM Labs, 30 Jan. 2025), p. 2.

² TRM Labs, p. 3.

³ *ibid*, p. 4.

⁴ *ibid*, p. 3.

⁵ Bondy, Meeka, and John F. Delaney. “AI-Generated Deepfakes and the Emerging Legal Landscape.” (Perkins Coie, 15 Apr. 2024), p. 2.

⁶ Nerantzi, Elina, and Giovanni Sartor. “‘Hard AI Crime’: The Deterrence Turn.” *Oxford Journal of Legal Studies*, vol. 44, no. 3, (Autumn 2024), p. 674.

responsibility gap,” as “no agent (human or artificial) can be legitimately punished for this outcome.”¹

B) Jurisdictional and Territorial Challenges

The cross-border nature of AI crimes presents significant jurisdictional and territorial challenges for criminal law enforcement. Unlike traditional crimes that typically occur within defined geographical boundaries, AI-enabled criminal activities often transcend national borders, operating in digital spaces that exist beyond conventional jurisdictional frameworks. As Claypoole observes, “Passing a law to ‘restrict artificial intelligence’ is a dangerous exercise under current definitions,”² highlighting the difficulty in crafting legislation that can effectively address the global nature of AI crimes.

The challenge is further complicated by the distributed nature of AI systems themselves. An AI-enabled criminal operation might involve algorithms developed in one country, data processed in another, and harmful effects manifesting in multiple jurisdictions simultaneously. This distribution of elements across borders raises complex questions about which nation’s laws apply and which authorities have jurisdiction to investigate and prosecute such crimes. As noted by Nanos, “a reliable legal framework is indispensable, particularly concerning issues of liability when unintended events occur due to AI technology.”³

International cooperation becomes essential in addressing these jurisdictional challenges, yet existing frameworks for mutual legal assistance were largely designed for a pre-digital era and struggle to keep pace with the rapid evolution of AI technologies. The need for harmonized approaches to AI crimes is evident, but achieving consensus on definitions, standards, and enforcement mechanisms across diverse legal systems remains a significant hurdle. Without effective international cooperation, jurisdictional gaps may be exploited by criminals, creating safe havens from which AI-enabled criminal activities can be conducted with relative impunity.

In reality, given the inherently global nature of AI development and deployment, enhanced international cooperation is indispensable.⁴ The jurisdictional challenges highlighted earlier underscore the need for international agreements, treaties, or conventions to harmonize legal approaches to AI criminality, facilitate cross-border investigations, and establish mechanisms for

¹Nerantzi and Sartor, op.cit. p. 673.

²Claypoole, op.cit. p. 1.

³ Nanos, op.cit. p. 1.

⁴See, e.g., Anne-Marie Slaughter, *A New World Order* (Princeton University Press, 2004) (discussing the role of international cooperation in a globalized world).

mutual legal assistance.¹ International bodies could play a key role in developing common standards, sharing best practices, and fostering dialogue on the complex legal and ethical issues raised by AI.² Without such cooperation, there is a risk of creating legal havens for irresponsible AI development or a patchwork of conflicting national laws that impede both innovation and effective redress for AI-perpetrated harms.³ The path forward requires a concerted, collaborative effort to ensure that the evolution of criminal law and its associated frameworks keeps pace with the transformative potential and inherent risks of autonomous artificial intelligence, safeguarding societal values without unduly stifling technological progress.⁴

Section II: The Criminal Responsibility Gap

A) Traditional Criminal Law Requirements

Traditional criminal law is founded on two fundamental elements: *actus reus* (the guilty act) and *mens rea* (the guilty mind). These elements form the bedrock of criminal culpability in most legal systems worldwide. The *actus reus* requirement ensures that criminal liability attaches only to voluntary actions or culpable omissions, while *mens rea* requires that the defendant possessed a certain mental state—whether intent, knowledge, recklessness, or negligence—in relation to the prohibited conduct or result.

These traditional requirements present significant challenges when applied to AI systems. As Nanos explains, “Future criminal law must be prepared and probably adjusted effectively tackle any questions concerning criminal liability of artificial intelligence.”⁵ The difficulty arises because AI systems do not possess mental states in the human sense, raising fundamental questions about whether they can satisfy the *mens rea* requirement for criminal liability.

Moreover, the *actus reus* requirement presupposes a level of agency and voluntariness that is difficult to attribute to AI systems. While an AI system may perform actions that cause harm, determining whether these actions constitute “voluntary” conduct in the legal sense is problematic. The autonomous nature of

¹See, e.g., United Nations, *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*, A/HRC/23/40 (Apr. 17, 2013) (discussing international norms for online content).

²See, e.g., OECD Principles on Artificial Intelligence (2019) (an example of international guidelines for AI).

³See, e.g., Tim Wu, *The Master Switch: The Rise and Fall of Information Empires* (Knopf, 2010) (discussing the risks of fragmented regulatory approaches).

⁴See, e.g., Luciano Floridi, *The Fourth Revolution: How the Infosphere is Reshaping Human Reality* (Oxford University Press, 2014) (discussing the transformative impact of information technologies).

⁵ Nanos, *op.cit.*p. 1.

advanced AI systems further complicates this analysis, as their actions may result from complex learning processes rather than direct programming or human control.

B) The “Hard AI Crime” Problem

The concept of “Hard AI Crime” represents one of the most challenging frontiers in the intersection of AI and criminal law. Nerantzi and Sartor define this as occurring when “the harmful consequence brought about by such an AI agent—an agent acting as a utility maximiser—was not intended by any of the humans who contributed to the agent’s design and deployment.”¹ In such scenarios, an AI system autonomously engages in conduct that would constitute a crime if performed by a human, yet no human involved in its creation or operation possessed the requisite criminal intent.

This creates what the authors term a “criminal responsibility gap,” as “no agent (human or artificial) can be legitimately punished for this outcome.”² The gap emerges because the AI system itself lacks the legal personhood and moral agency traditionally required for criminal responsibility, while the humans involved lack the mens rea necessary for criminal liability.

Nerantzi and Sartor propose shifting the discussion “from blame to deterrence” by developing an “AI deterrence paradigm” inspired by economic theories of crime.³ They suggest that even if AI systems cannot be meaningfully blamed in the moral sense, they can be effectively deterred since they internalize sanctions as costs. This approach reconceptualizes the AI system as a *machinaeconomica* that responds to incentives, even if it lacks moral understanding.

C) Liability Models for AI Crimes

Various liability models have been proposed to address the criminal responsibility gap in AI crimes. Direct liability approaches focus on holding developers, users, or companies accountable for harmful AI behaviors. As noted by Nanos, “a well-developed set of rules should not only concern private law... when such technology causes harm or even death to human beings, the question of a criminal deed arises.”⁴ Under this model, criminal liability might attach to developers who negligently design AI systems without adequate safeguards, or to users who deploy AI systems in contexts where foreseeable harm could result.

Vicarious liability frameworks offer another approach, drawing parallels to corporate criminal liability. Just as corporations can be held criminally liable

¹Nerantzi and Sartor, op.cit.p. 674.

²ibid, p. 673.

³ibid.

⁴ Nanos, op.cit.p. 1.

for the actions of their employees under certain circumstances, similar principles might be applied to hold organizations responsible for the actions of their AI systems. This approach recognizes that while AI systems lack legal personhood, they operate as agents of the organizations that deploy them.

Strict liability considerations represent a third model, potentially imposing criminal liability without requiring proof of mens rea for certain high-risk AI applications. As observed in research on AI liability frameworks, “This would allow regulators to apply more stringent liability rules to high-risk AI applications, such as those used in healthcare or criminal justice.”¹ While strict liability runs counter to traditional criminal law principles that require culpable mental states, it may be justified in contexts where AI systems pose significant risks to public safety or fundamental rights.

Section III: AI Risk Assessment in Criminal Justice

A) Current Applications in Judicial Determinations

AI risk assessment tools have become increasingly prevalent in various stages of the criminal justice process. As Sanger notes, “The use by the courts of risk assessment programs is one of the principal incursions of Artificial Intelligence (AI) into criminal law judicial determinations.”² These tools are employed in pretrial decisions, sentencing, and parole determinations, often with significant consequences for the individuals subject to their assessments.

In the pretrial context, risk assessment algorithms evaluate factors such as criminal history, age, and community ties to predict the likelihood that a defendant will fail to appear for court proceedings or commit new offenses if released. These predictions inform judicial decisions about pretrial detention, bail amounts, and release conditions. Similarly, at the sentencing stage, risk assessment tools provide judges with recidivism predictions that may influence incarceration decisions. As Sanger observes, “The decision can change lives forever—not just of the person incarcerated but their family, loved ones and community.”³

Parole decision support systems represent another significant application of AI in criminal justice. These systems analyze various factors to predict an inmate’s likelihood of reoffending if released, potentially influencing parole board decisions about whether to grant early release. The widespread adoption

¹ Daily Journal. “Legal liability concerns in artificial intelligence: What you need to know.” Daily Journal, (11 Sept. 2024), p. 1.

²RobertSanger. “Artificial Intelligence and Criminal Law.” (The Colleges of Law, 12 Jan. 2024), p. 1.

³ibid, p. 2.

of these tools reflects a belief that data-driven approaches can enhance the objectivity and consistency of criminal justice decisions.

B) Validity and Bias Concerns

Despite their growing prevalence, AI risk assessment tools face serious questions regarding their validity and potential for bias. Sanger argues that “the science does not bear out” the assumption that these tools are more objective than human judges.¹ He notes that “so-called ‘validated risk assessment tool[s]’ and ‘risk assessment instrument[s]’” may give a false impression of scientific rigor through their clinical-sounding terminology.²

Research from Boston University highlights how algorithms built upon biased data can “replicate or even amplify that bias when they spit out results.”³ This is particularly concerning in the criminal justice context, where historical data reflects systemic biases against marginalized communities. As the research notes, one risk assessment system “guessed wrong about two times as often for Black people than for white people: it mislabeled twice as many Black people who didn’t reoffend as being at high risk for doing so.”⁴

Transparency and explainability present additional challenges. Many risk assessment algorithms operate as “black boxes,” making it difficult for defendants, attorneys, and even judges to understand how they reach their conclusions. This opacity undermines due process principles, as individuals cannot effectively challenge assessments they do not understand. As noted in research on AI in criminal justice, “One of the most contentious issues with AI in criminal justice is its lack of transparency.”⁵

C) Human vs. AI Decision-Making

The relationship between human and algorithmic decision-making in criminal justice raises complex questions about the appropriate division of labor. While AI systems can process vast amounts of data and identify patterns that might escape human notice, they lack the contextual understanding, empathy, and moral reasoning capabilities that human judges bring to their decisions. As Sanger observes, “unaided by a supercomputer, the human brain necessarily

¹ibid.

²ibid.

³ Boston University. “Algorithms Were Supposed to Reduce Bias in Criminal Justice—Do They?” (The Brink, Boston University, 23 Feb. 2023), p. 2.

⁴ Boston University, p. 2.

⁵ The Legal Wire. “AI in Criminal Justice: The Bias and Accountability Dilemma.” (The Legal Wire, 4 Feb. 2025), p. 1.

reverts to shortcuts which are based on prior observations, generalizations, stereotypes, and biases.”¹

Hybrid approaches that combine human and algorithmic assessments may offer a path forward. Under such approaches, AI tools provide information and recommendations, but human decision-makers retain ultimate authority and responsibility. This preserves human judgment while potentially enhancing it with data-driven insights. However, research suggests that humans may exhibit “automation bias,” deferring to algorithmic recommendations even when they have reason to be skeptical.

Safeguards and oversight mechanisms are essential regardless of the approach adopted. These might include regular audits of algorithmic systems for bias and accuracy, requirements for explainability and transparency, and meaningful opportunities for individuals to challenge algorithmic assessments. As Boston University research suggests, “Really delivering on the promise of algorithms in the criminal justice system—the promise that they make the process more uniform and less biased—will require a radical reimagining of their use.”²

Section IV: Emerging Legal Frameworks for AI-Specific Crimes

A) Deepfakes and Digital Impersonation

The proliferation of deepfakes—synthetic media in which a person’s likeness is replaced with someone else’s using AI techniques—has prompted legislative responses at both state and federal levels. As Bondy and Delaney note, “In the wake of these and other high-profile incidents, state and federal legislators are scrambling to address the unauthorized and nonconsensual use of deepfakes.”³ These legislative efforts have focused on three primary areas: nonconsensual sexually explicit deepfakes, election-related deepfakes, and right of publicity concerns.

At the state level, several jurisdictions have enacted laws specifically targeting nonconsensual sexually explicit deepfakes. Some states, including “Georgia, Hawaii, New York, and Virginia,” have amended existing “revenge porn” laws to expressly include deepfakes.⁴ Others, such as “Florida, Minnesota, and Texas,” have created standalone statutes addressing this issue.⁵ These laws vary in their definitions, intent requirements, exemptions, and available remedies, creating a patchwork of protections across different jurisdictions.

¹ Sanger, *op.cit.* p. 2.

² Boston University, *op.cit.*p. 3.

³Bondy and Delaney, *op.cit.*p. 2.

⁴*ibid.*

⁵*ibid.*

Federal legislative proposals have also emerged, though none have yet been enacted into law. The “Disrupt Explicit Forged Images and Non-Consensual Edits Act of 2024 (DEFIANCE Act)” would create a civil remedy for certain nonconsensual sexually explicit deepfakes.¹ Similarly, the “Preventing Deepfakes of Intimate Images Act” would establish both civil and criminal liability for such content.² These proposals reflect growing recognition of the harm caused by deepfakes and the need for federal standards to address them.

B) Algorithmic Bias and Discrimination

The issue of algorithmic bias in criminal justice has attracted increasing attention from lawmakers and regulators. As research from Boston University highlights, “In theory, if the predictive algorithm is less biased than the decision-maker, that should lead to less incarceration of Black and Indigenous and other politically marginalized people. But algorithms can discriminate.”³ This recognition has prompted efforts to develop legal frameworks addressing algorithmic discrimination.

One approach focuses on transparency and accountability requirements. These may include mandates for algorithmic impact assessments, regular audits for bias, and disclosure of the data and methodologies used in developing algorithmic systems. As noted in research on AI in criminal justice, “Algorithms used in criminal justice should be transparent, with clear explanations of how decisions are reached.”⁴

Another approach emphasizes community involvement in the development and oversight of algorithmic systems. Boston University research argues that “the very paradigm that governs if and how we use these algorithms is quite technocratic and not very diverse” and that “the turn to algorithms is inconsistent with a racial justice project, because of the way in which they maintain the marginalization of these same communities.”⁵ Legal frameworks that require meaningful input from affected communities may help address these concerns.

C) Autonomous Systems Liability

The question of liability for autonomous AI systems presents particularly complex challenges for legal frameworks. As Nanos observes, “certain AI applications, such as autonomous driving, possess inherent risks that could

¹ibid. p. 3.

²ibid.

³ Boston University, op.cit. p. 1.

⁴ Cloud Tweaks. “The Ethics of AI Decision-Making in the Criminal Justice System.” (Cloud Tweaks, 27 Sept. 2024), p. 1.

⁵ Boston University, op. cit. p. 2.

potentially infringe upon individuals' protected rights, and even cause physical harm and death to human beings.”¹ These risks necessitate clear liability rules that balance innovation with accountability.

Regulatory approaches to autonomous AI vary across jurisdictions and contexts. Some frameworks adopt a risk-based approach, imposing stricter requirements on high-risk AI applications while allowing greater flexibility for lower-risk systems. Others focus on specific sectors, such as autonomous vehicles or healthcare AI, developing tailored regulations that address the unique challenges in each domain.

The distinction between criminal and civil liability remains significant in this context. While civil liability mechanisms may provide compensation for harm caused by autonomous systems, they may not adequately address the societal interest in deterring dangerous AI deployments. As research on AI liability notes, “Product liability and negligence frameworks struggle to address the complexities of AI-driven systems.”² Criminal liability may be necessary in cases involving gross negligence or reckless disregard for foreseeable risks.

Given the doctrinal impediments to assigning direct criminal liability to autonomous AI entities under existing legal paradigms, the critical question of accountability for harms caused by such systems necessitates an exploration of alternative and adapted theoretical frameworks.³ The lacuna in direct AI culpability does not absolve the legal system of the responsibility to identify a locus of liability; rather, it compels a nuanced examination of how responsibility can be attributed when autonomous technologies are implicated in conduct that would, if perpetrated by a human, attract criminal sanction. Several principal avenues of liability warrant meticulous consideration, each presenting its own set of advantages, limitations, and jurisprudential challenges. These include various forms of vicarious liability, potentially extending to manufacturers, programmers, deployers, or owners of AI systems;⁴ the exploration of organizational liability, drawing parallels with corporate criminal responsibility;⁵ and the more radical proposition of developing novel legal

¹ Nanos, op.cit. p. 1.

² YRIS. “Navigating Liability In Autonomous Robots: Legal And Ethical Challenges in Manufacturing and Military Applications.” Yale Review of International Studies, (6 Mar. 2025), p. 1.

³ See, e.g., Ryan Abbott, *The Reasonable Robot: Artificial Intelligence and the Law* (Cambridge University Press, 2020) (examining various legal frameworks for AI).

⁴ See, e.g., Matthew U. Scherer, *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies*, 29 Harv. J.L. & Tech. 353 (2016) (discussing liability for AI systems).

⁵ See, e.g., *New York Central & Hudson River Railroad Co. v. United States*, 212 U.S. 481 (1909) (a key case in establishing corporate criminal liability).

categories or statuses specifically tailored to advanced AI, which might entail a sui generis liability regime.¹

International perspectives offer valuable insights for developing effective liability frameworks. The European Union's proposed AI Act, for instance, adopts a risk-based approach that imposes varying levels of obligations based on an AI system's potential for harm. Such approaches recognize that one-size-fits-all solutions are unlikely to be effective given the diverse applications and risk profiles of AI technologies.

Conclusion:

The rapid advancement of artificial intelligence has created unprecedented challenges for criminal law frameworks worldwide. As this article has demonstrated, these challenges span multiple dimensions: from categorizing novel forms of AI-enabled criminal activity, to addressing jurisdictional complexities, confronting the criminal responsibility gap, evaluating AI risk assessment tools, and developing targeted legal responses to AI-specific crimes. The legal system's ability to effectively address these challenges will significantly impact both the development of AI technologies and the protection of fundamental societal values.

The criminal responsibility gap represents perhaps the most profound challenge to traditional criminal law principles. When autonomous AI systems engage in harmful behaviors without clear human intent or negligence, existing liability frameworks struggle to provide adequate responses. As Nerantzi and Sartor argue, shifting from a blame-focused approach to a deterrence paradigm may offer a promising path forward, reconceptualizing AI systems as economic actors that respond to incentives even if they lack moral agency.² However, this approach requires careful implementation to ensure that it does not undermine the moral foundations of criminal law or create perverse incentives for AI developers and deployers.

The use of AI within the criminal justice system itself presents additional concerns regarding validity, bias, and appropriate governance. As Sanger cautions, the clinical-sounding terminology of "validated risk assessment tool[s]" may create a false impression of scientific rigor.³ The evidence suggests that these tools can replicate or even amplify existing biases, particularly against marginalized communities. Addressing these issues requires not only technical

¹See, e.g., Mireille Hildebrandt, *Smart Technologies and the End(s) of Law: Novel Entanglements of Law and Technology* (Edward Elgar Publishing, 2015) (exploring the transformative impact of technology on law).

²Nerantzi and Sartor, op.cit. p. 673.

³ Sanger, op.cit. p. 2.

solutions but also procedural safeguards and meaningful community involvement in the development and oversight of these systems.

Emerging legal frameworks for AI-specific crimes such as deepfakes show promise but remain fragmented and incomplete. The patchwork of state laws addressing nonconsensual sexually explicit deepfakes, for instance, creates inconsistent protections across jurisdictions. Federal legislation could provide more uniform standards, but must be carefully crafted to balance free expression concerns with the prevention of serious harms. Similarly, approaches to algorithmic bias and autonomous systems liability are still evolving, with significant variations across different legal systems.

Based on this analysis, several recommendations emerge for legal and policy reforms:

First, criminal law frameworks should be updated to explicitly address the unique challenges posed by autonomous AI systems. This may include developing new liability models that recognize the distributed nature of AI development and deployment, establishing clear standards of care for AI developers and users, and potentially creating new categories of offenses specifically tailored to AI-enabled harms. As Claypoole suggests, breaking AI into “functional categories that raise similar issues for the people affected by the technology in that category will simplify effective rulemaking.”¹

Second, international cooperation should be strengthened to address the cross-border nature of AI crimes. This includes harmonizing definitions and standards where possible, enhancing mechanisms for mutual legal assistance in investigations involving AI systems, and developing protocols for sharing technical expertise across jurisdictions. Without such cooperation, jurisdictional gaps will continue to be exploited by malicious actors.

Third, the use of AI within the criminal justice system should be subject to rigorous oversight and governance. This includes mandatory transparency requirements, regular audits for bias and accuracy, meaningful opportunities for defendants to challenge algorithmic assessments, and community involvement in the development and implementation of these systems. As Boston University research emphasizes, “the turn to algorithms is inconsistent with a racial justice project” unless marginalized communities are centered in decision-making processes.²

Fourth, targeted legal responses to AI-specific crimes should be developed at both state and federal levels. For deepfakes, this includes comprehensive legislation addressing not only nonconsensual sexually explicit

¹Claypoole,op.cit. p. 3.

² Boston University,op.cit. p. 2.

The Criminal Law Challenges in Confronting AI Crimes

content but also election-related manipulation and fraud. For algorithmic bias, this includes clear anti-discrimination standards that apply to AI systems and meaningful enforcement mechanisms. For autonomous systems, this includes liability frameworks that appropriately balance innovation with accountability.

Finally, interdisciplinary collaboration should be fostered between legal experts, technical specialists, ethicists, and affected communities. The complexity of AI technologies and their societal implications demands diverse perspectives and expertise. Legal professionals must develop sufficient technical literacy to understand AI capabilities and limitations, while technical experts must appreciate the legal and ethical dimensions of their work.

As society continues to integrate increasingly autonomous AI systems into critical functions, the criminal justice system must evolve to address the novel challenges these technologies present. This evolution requires not only technical understanding but also careful consideration of the ethical, social, and legal implications of attributing criminal liability in an age of artificial intelligence. By thoughtfully adapting criminal law frameworks to the realities of AI while preserving fundamental principles of justice, fairness, and accountability, we can harness the benefits of these technologies while mitigating their potential harms.

The path forward will not be straightforward, and will require ongoing reassessment as AI technologies continue to evolve. However, by proactively addressing the criminal law challenges in confronting AI crimes, rather than reacting to harms after they occur, we can help ensure that the development and deployment of AI technologies aligns with our collective values and serves the public good. The stakes are significant, but with thoughtful analysis, interdisciplinary collaboration, and a commitment to both innovation and protection, we can navigate this complex terrain effectively.

List of sources and references:

1. Abbott, Ryan. *The Reasonable Robot: Artificial Intelligence and the Law*. Cambridge University Press, 2020.
2. Bondy, Meeka, and John F. Delaney. "AI-Generated Deepfakes and the Emerging Legal Landscape." *Perkins Coie*, 15 Apr. 2024.
3. Boston University. "Algorithms Were Supposed to Reduce Bias in Criminal Justice—Do They?" *The Brink*, Boston University, 23 Feb. 2023.
4. Claypoole, Theodore Franklin. "AI Classifications for Law and Regulation." *American Bar Association*, 15 Sept. 2023.
5. Cloud Tweaks. "The Ethics of AI Decision-Making in the Criminal Justice System." *Cloud Tweaks*, 27 Sept. 2024.
6. Daily Journal. "Legal liability concerns in artificial intelligence: What you need to know." *Daily Journal*, 11 Sept. 2024.
7. Floridi, Luciano. *The Fourth Revolution: How the Infosphere is Reshaping Human Reality*. Oxford University Press, 2014.

8. Hildebrandt, Mireille. *Smart Technologies and the End(s) of Law: Novel Entanglements of Law and Technology*. Edward Elgar Publishing, 2015.
9. Nanos, Andreas. "Criminal Liability of Artificial Intelligence." *Charles University in Prague Faculty of Law Research Paper No. 2023/III/3*, 4 Nov. 2023.
10. Nerantzi, Elina, and Giovanni Sartor. "'Hard AI Crime': The Deterrence Turn." *Oxford Journal of Legal Studies*, vol. 44, no. 3, Autumn 2024.
11. *New York Central & Hudson River Railroad Co. v. United States*, 212 U.S. 481 (1909).
12. OECD. *Principles on Artificial Intelligence*. 2019.
13. Sanger, Robert. "Artificial Intelligence and Criminal Law." *The Colleges of Law*, 12 Jan. 2024.
14. Scherer, Matthew U. "Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies." 29 *Harv. J.L. & Tech.* 353 (2016).
15. Slaughter, Anne-Marie. *A New World Order*. Princeton University Press, 2004.
16. The Legal Wire. "AI in Criminal Justice: The Bias and Accountability Dilemma." *The Legal Wire*, 4 Feb. 2025.
17. TRM Labs. "The Rise of AI-Enabled Crime: Exploring the evolution, risks, and responses to AI-powered criminal enterprises." *TRM Labs*, 30 Jan. 2025.
18. United Nations. *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*, A/HRC/23/40, 17 Apr. 2013.
19. Wu, Tim. *The Master Switch: The Rise and Fall of Information Empires*. Knopf, 2010.
20. YRIS. "Navigating Liability In Autonomous Robots: Legal And Ethical Challenges in Manufacturing and Military Applications." *Yale Review of International Studies*, 6 Mar. 2025.