

Criminal Liability of the Non-Human Actor in the Age of Artificial Intelligence: A Theoretical Approach to the Responsibility of Autonomous Systems

المسؤولية الجنائية للفاعل غير البشري في عصر الذكاء الاصطناعي: مقارنة
نظرية لمسؤولية الأنظمة المستقلة

Ferhi Rabiaa¹

¹Faculty of Law and Political Sciences, University Echahid Cheikh Larbi Tebessi (Algeria),
rabiaa.ferhi@univ-tebessa.dz

Received: 21/07/2025

Accepted: 13/11/2025

Published: December/2025

الملخص:

إن ظهور كيانات حديثة في عصر الذكاء الاصطناعي يعد أبرزها الأنظمة المستقلة التي شهدت طفرة في قدراتها، أدت إلى امتلاكها قدرًا متزايدًا من الاستقلالية في اتخاذ القرار، هو أمر أفضى في بعض الحالات إلى انتهاك قواعد القانون الجنائي، سواء بارتكاب أفعال جرمية مباشرة أو التسبب بها. في هذا السياق، برزت إشكالية محورية تتعلق بمدى قابلية هذه الأنظمة، باعتبارها فاعلاً غير بشري للمساءلة الجنائية وفقاً للمبادئ التقليدية للمسؤولية الجزائية، لا سيما ما يتصل بالركن المعنوي. كما طُرحت تساؤلات حول الأساس القانوني لإسناد الجريمة، بين إمكانية تأسيس مسؤولية مباشرة للأنظمة المستقلة في طرح آثار العديد من المسائل أهمها الاعتراف بالشخصية القانونية لهذه الأنظمة، وحصر المسؤولية في الإطار البشري التقليدي، أو إمكانية مساءلتها معاً.

الكلمات المفتاحية: الفاعل غير البشري؛ جرائم الأنظمة المستقلة؛ المسؤولية الجزائية؛ الذكاء الاصطناعي.

Abstract:

The rise of new entities in the Artificial Intelligence era, especially autonomous systems, has led to unprecedented independence in decision-making. This development has, at times, resulted in violations of criminal law, either directly or indirectly. A core issue arises regarding the criminal liability of these non-human actors under traditional principles. The focus is particularly on the mental element (*Mens Rea*) in criminal responsibility. Debates revolve around whether autonomous systems can bear direct liability, requiring legal personality recognition. Alternatively, liability may remain strictly human-centered assigned to developers, programmers, or users. Another proposition suggests shared liability between the system and human actors. These questions reflect a deeper legal dilemma in regulating AI-related crimes.

Key words: Non-human Actor; Crimes of Autonomous Systems; Criminal Liability; Artificial Intelligence

Introduction:

Criminal liability is more specific than civil liability, according to the construction of the criminal theory under traditional principles, which requires many elements to hold criminals responsible, to achieve justice through human-designed criminal law, which governs human conduct and its violations.

Over the last two decades, criminal law has been required to supply adequate solutions for the commission of criminal offenses, as other entities are not recognized as legal personalities in the criminal law, and artificial intelligence (AI) entities are pivotal topics in this domain.

The use of AI is growing in many ways within the modern world. Nowadays, this technology plays a crucial role in our daily lives, from personal assistants to self-driving cars. As a disruptive technology, AI is being widely used in every aspect of society, drastically changing the world we live in. Those who have an advantage over AI will be able to gain the initiative in the competition to control the future. As a consequence, the advancement of AI was formally recognized as a matter of national strategic importance.

The rapid rise of AI has gone beyond mere technological advancement, with the emergence of crimes directly linked to AI programs and autonomous systems, so-called *Sapiens Machina*¹, that not only learn and evolve but sometimes seem to act with independent judgment or purpose. This development compels us to rethink how legal responsibility is understood, especially in criminal law. As machines become more autonomous and capable of making complex decisions on their own, the idea of a "perpetrator" is no longer something that applies only to humans.

From this point arises the significance of the present study; the importance of this study arises because criminal liability traditionally requires the availability of specific elements to hold perpetrators accountable for their actions. However, this raises critical questions about whether conventional criminal law is truly equipped to deal with emerging crimes committed by artificial intelligence entities today.

Moreover, there is growing concern about the fate of traditional legal frameworks in the face of rapid advancements in smart robotics. Given the current uncertainties surrounding the future scope of AI development and its potential applications, it is premature to adopt definitive or prejudiced positions regarding the attribution of criminal liability to such systems. It would be an oversimplification to assume that algorithms, which always make decisions within a set framework, imply that only the human programmers should be held accountable. While AI may not yet rival human intelligence, the future promises groundbreaking leaps that could grant these entities increasing levels of autonomy, posing new and complex legal and ethical challenges that can no longer be ignored.

Accordingly, this research paper seeks to address the main question: **What is the theoretical basis for assigning criminal liability to autonomous artificial intelligence systems?**

To answer this question, a descriptive approach is adopted to introduce the concept of artificial intelligence and trace the evolution of its technologies. In addition, an analytical framework is employed to critically examine the legal and ethical implications of holding AI

systems criminally accountable. This dual approach aims to provide a comprehensive understanding of the challenges posed by emerging AI technologies in the field of criminal law.

To achieve its purposes, this research is divided into two main sections:

- The Concept of AI Autonomous Systems in the Criminal Law Context
- Towards a Theoretical Conceptualization of the Non-Human Actor

1-The Concept of AI Autonomous Systems in the Criminal Law Context

The rapid rise of AI has opened the door to a new world of opportunities, reshaping industries and influencing the way our societies function. The relationship between criminal law and AI systems has also attracted attention, primarily in terms of how AI affects the functioning of the criminal justice system through the use of AI tools to forecast the likelihood of criminal activity or to identify existing or potential future crime victims².

Yet, this technological leap also puts pressure on long-standing legal principles, especially those related to criminal responsibility. As AI systems grow more autonomous and complex, it becomes increasingly important to rethink and adapt our legal frameworks to meet these new realities. This need becomes even more pressing with the emergence of Artificial General Intelligence systems capable of performing intellectual tasks at or beyond human level. Unlike narrow AI, which operates within predefined boundaries, this technology could act with a degree of independence and decision-making power that challenges existing notions of control, accountability, and liability. Therefore, legal systems must evolve to address the unprecedented ethical and legal implications posed by such advanced forms of intelligence.

Determining who should be held accountable when machines make decisions is not a simple task, and as AI continues to evolve, our legal and ethical responses must evolve accordingly³.

1.1- The Nexus between the Development of AI and Autonomous Systems

The concepts of AI and autonomous systems are related and complementary; the idea of AI and autonomous systems can be traced back thousands of years, when ancient philosophers explored the mysteries of life and death. Long ago, inventors designed mechanical devices known as "automatons" machines that could move on their own without human assistance. The term "automaton" is derived from ancient Greek and means "self-operating" or "acting independently".

One of the earliest known examples of this technology dates back to around 400 BC and involves a mechanical bird built by a companion of the philosopher Plato. Centuries later, around 1495, Leonardo da Vinci created one of the most well-known automatons in history⁴.

The rise of the term AI was in the 20th century, as the roots of its development date back to the work of Vannevar Bush, the American scientist, "As We May Think", which proposed a system to improve the ability of human understanding and knowledge. Also, the British mathematician Alan Turing, who proposed the idea that machines can simulate human being behavior and intelligence, like playing chess in his published article about Computing Machinery and Intelligence⁵. He outlined a method of building intelligent machines and, more

Criminal Liability of the Non-Human Actor in the Age of Artificial Intelligence: A Theoretical Approach to the Responsibility of Autonomous Systems
Ferhi Rabiaa

importantly, proposed a way to evaluate their intelligence. This method, now known as the Turing Test, remains a standard for assessing artificial intelligence. According to the test, if a person communicates with both a human and a machine without being able to tell which is which, the machine can be considered intelligent⁶.

The use of the term "Artificial Intelligence" was coined in 1956 during the gathering of the scientific community at the Dartmouth Conference, when McCarthy had machine learning in mind⁷. In this era, machines could not reason and make rational and autonomous choices. Most of what is lately labelled intelligence is computer processing⁸. During the two decades after the first use of the term AI. Between 1964 and 1966, this field witnessed the development of the ELIZA computer program, the General Problem Solver, and other pioneering systems that contributed substantially to what we know today as artificial intelligence.

The competition to build technology influenced the development of AI. Since Japan's Ministry of Economy, Trade and Industry had been investing large amounts of money to develop what was known in 1981 as the "AI computer", as a part of the 5th Generation Computer Project. the governments of the United Kingdom and the United States sought to simulate these procedures by increasing their funding for research in information technology⁹.

The third wave of the AI evolution began in 1993. In the decades following series of technological milestones laid the groundwork for the rapid advancement of AI. Beginning with the U.S. government's strategic vision to build a national information infrastructure, the momentum of innovation accelerated as breakthroughs in computing power, natural language processing, neural networks, and deep learning emerged. Major tech companies and academic institutions across the globe invested heavily in AI research, leading to the development of intelligent systems capable of outperforming humans in specific tasks, simulating cognitive processes, and transforming user experiences across digital platforms. These collective efforts established a solid foundation for the explosive growth of AI as a central force in technological progress¹⁰.

During the last years, the use of AI has influenced the transition from basic automation to advanced autonomous systems, which has, in many ways, coincided with the later stages of development in computer science and robotics¹¹. In recent years, AI has quietly but profoundly evolved, thanks to machine learning, especially deep learning. A key breakthrough came in 2012, when Geoff Hinton's team won the ImageNet competition, proving the power of neural networks trained through backpropagation. Though the idea dated back to 1986, it only became practical with modern computing power and big data. This moment marked a turning point, pushing AI and autonomous systems into a new era¹².

Since then, progress has accelerated rapidly. Deep learning now powers everything from voice assistants to self-driving cars, and AI systems are increasingly capable of making complex decisions with minimal human input. Today, AI is not just a research field; it is a core part of everyday technologies and a driving force behind autonomous systems shaping industries, defense, and even legal frameworks. What began as a theoretical idea has become an essential tool transforming how we live, work, and govern.

1.2- Contemporary Models of Autonomous Systems Crimes that Sparked the Liability Debate

After the great jump in the technologies' domain, some incidents raised the issue of the possibility of accounting the autonomous systems as non-human actors, especially under the gaps of legislations and the lack of a clear theory towards a legal framework to establish a new era of criminal responsibility theory. these are some relevant cases from the last years.

1.2.1-Uber Self-Driving Car Fatality: In March 2018, a fatal crash involving an autonomous Uber vehicle, it was the first pedestrian fatality involving a fully autonomous vehicle, which triggered both technical reforms within Uber's program and broader regulatory scrutiny of autonomous vehicle testing¹³.

Criminal prosecutors in Arizona dropped the case against Uber, meaning the company will not face criminal charges. However, this does not absolve Uber from all legal responsibility. The vehicle operator may still face criminal prosecution for manslaughter due to inattentiveness. In fact, it is unlikely that other Uber staff, such as engineers or testers, were found criminally liable because they committed a criminal offense¹⁴.

While Uber, as a legal person, could be criminally liable, it was not prosecuted in this case. Importantly, the vehicle itself cannot be held criminally liable, as it is not a legal entity. According to the traditional theory of crimes, criminalizing a machine is legally and logically invalid¹⁵.

1.2.2- Deepfake Related Offenses: Deepfake technology uses AI to create highly realistic but fake videos, audio, or images that are difficult to distinguish from real ones. Techniques like face swapping, voice cloning, and full-body synthesis allow users to mimic people convincingly. When used to harm someone's reputation, deepfakes become a form of cyber defamation¹⁶.

In 2019, a German branch of a British energy company fell victim to a deepfake voice scam in which criminals used AI-generated audio to impersonate the company's UK-based CEO. Using sophisticated voice-mimicking technology, the fraudsters impersonated an executive to deceive a senior manager into transferring €220,000 to a fictitious Hungarian supplier¹⁷.

This real-life case highlights the growing risks posed by AI technologies, particularly deepfake tools, in enabling sophisticated crimes without direct human involvement at the time of execution. It raises serious legal questions about liability, whether it lies with the user, the programmer, or the provider of the AI tool, emphasizing the urgent need to adapt the criminal responsibility framework to address non-human actors¹⁸.

1.2.3- Autonomous weapons systems: the world has been changing in response to the big wave of technology advancement, and so has the war. Humanity is at the dawn of a new era of warfare due to the AI's impact, with major consequences for global security¹⁹.

Autonomous weapon systems, known as killer robots, are defined as any weapon systems with autonomy in their critical functions²⁰, or any weapon systems that can select and attack targets without human intervention, so the weapon itself uses its sensors, software, and weaponry systems in addition to the human control capability²¹.

Criminal Liability of the Non-Human Actor in the Age of Artificial Intelligence: A Theoretical Approach to the Responsibility of Autonomous Systems
Ferhi Rabiaa

Autonomous weapon systems are endangering human rights and putting them at serious risk²². After the widespread use of these weapons in wars and conflicts, which has raised the issue of legal considerations of using these weapons, especially the criminal responsibility for these machines' conduct under various hypotheses based on the probability of recognition personhood for these machines or the complete assumption of human control over these devices²³.

According to this evolution of weaponry, the international humanitarian law is seeking to control and regulate the use of autonomous agents to ensure the respect of war's rules, customs, and ethics.

Lately, Discussions regarding these weapons have been revived with greater intensity than in the past, because of their use in many conflicts. the Russian-Ukrainian war, which has witnessed operations based on autonomous weapons from both parties, such as the Russian military's use of drone swarms that consist of numerous drones equipped by AI systems, making them act autonomously, or the Phantom automated turret system, which is used to defend Ukrainian military positions²⁴.

Not far, the war on Gaza also witnessed many situations involving the use of modern technological weapons, causing massive human and material losses for the Palestinians. In addition to the drones, the Israeli military used artificial intelligence-based systems to guide its strikes on the Gaza Strip, such as the "Gospel" (Habsora) and "Lavender"²⁵. These systems committed grave breaches against civilians due to the lack of accuracy in their targeting many times, which led them to be criticized by military and law experts, until know these weapons are committing mass slaughters against innocent people without any accountability.

2-Towards a Theoretical Conceptualization of the Non-Human Actor

If, in the distant future, AI gains the ability for practical reasoning and becomes an autonomous agent, the existing principles of criminal responsibility, regardless of their form at that time, would apply. There would be no need to create entirely new legal frameworks as long as these machines can make rational decisions.

However, during the extended transitional phase where AI systems operate with partial rather than full autonomy, temporary legal rules would be necessary. Currently, AI lacks both full and partial autonomous decision-making capabilities. In civil law, regulations like product liability already extend to AI, since it is still regarded merely as a tool controlled by human users²⁶.

2.1- Criminal Capacity of the Autonomous Systems in the Light of the Classical Criminal Theory

Law is essentially a system of rules that helps organize how people behave and interact within society, backed by the authority of the state when those rules are broken. At the heart of every legal system is the idea of the "person" because laws are made for people. As social beings, we naturally form relationships and live in communities, and to keep those relationships peaceful and orderly, we create rules everyone is expected to follow²⁷.

Capacity is the set of personal traits or psychological factors that must be present in a person in order for us to attribute the criminal act to him as having committed it with awareness and will, or it is the ability of a person to understand the nature of his actions and

Criminal Liability of the Non-Human Actor in the Age of Artificial Intelligence: A Theoretical Approach to the Responsibility of Autonomous Systems
Ferhi Rabiaa

assess their consequences. It is the suitability of the person who committed a crime to be held accountable for it.

Most criminal legislations include provisions that explicitly require the presence of the elements of perception and will as prerequisites for capacity that establish criminal liability, while other legislations include them in the grounds for exclusion of criminal responsibility²⁸.

In criminal law, blame and responsibility are typically linked to the defendant's ability to act with rationality and voluntary control. In cases like murder or robbery, the legal system operates on the premise that individuals understand the nature and consequences of their actions and are capable of choosing otherwise. This assumption aligns with retributive justice theories, which seek not only to penalize the offender but also to hold them morally accountable for their wrongful conduct²⁹.

In the same context, the question raised is: what if an AI autonomous system is the actor who commits a crime, regardless of its gravity? According to the traditional theory of responsibility, robots cannot be held criminally responsible because they lack legal personhood and the capacity for free will. In essence, robots are not seen as true agents capable of independent action, and therefore, they fall outside the scope of criminal liability.

As a result, when a robot causes harm, the focus of responsibility typically shifts to the human actors behind it, such as the manufacturer, the programmer, or the operator. This reality makes it essential to carefully distinguish between the legal accountability of these human parties and the theoretical question of whether a robot itself could ever bear responsibility for its actions³⁰.

Given the use of AI in the commission of crimes, legal scholarship faces significant challenges in establishing criminal intent, largely due to the complexity and opacity of AI systems' operational processes, which makes it difficult to confirm the user's intent to cause harm.

Conversely, negligent acts are considered the most likely form of liability in this context; however, proving negligence is equally complex, as it requires assessing the offender's ability, under real-life circumstances, to foresee the potential harm, even if they did not intend it³¹. This raises further challenges when differentiating negligence from intent in specific crimes, particularly those where reckless disregard for safety may border on willful blindness. In such cases, the line between failing to act with due care and consciously ignoring known risks becomes blurred, complicating the legal analysis of culpability.

For certain offenses, such as manslaughter, negligent behavior may suffice for liability, whereas for crimes requiring specific intent—like murder or premeditated harm—the absence of deliberate purpose excludes negligence as a basis for prosecution. This distinction is crucial in both criminal law and emerging areas like autonomous technologies, where assigning intent to non-human actors or supervising individuals involves complex normative and factual considerations.

2.2- Models For The Criminal Liability of AI Autonomous Systems

Theorizing the liability of autonomous systems comes with challenges; it is particularly complex to adapt this responsibility to modern Constitutions, especially the concurrence between the living and the digital. Indeed, it is necessary to assess whether a machine can

commit crimes or is just a tool, to determine the extent to which it can act in concurrence with human agents, and to clarify how much responsibility, if any, can be attributed to it³².

The application of the traditional theory of criminal responsibility sets some ideas to ensure accountability for perpetrators, regardless of who or what they are, humans or machines. For this reason, this paper focuses on the theoretical hypotheses that involve some rules fit with AI crimes, especially the autonomous systems as mentioned above, to fulfill criminal justice for the community and victims.

2.2.1 The Perpetration-by-Another Liability Model

In some situations involving strict liability offences, the legal system goes with the accountability of perpetrators-by-another model, either in some classical crimes or in modern ones, such as media crimes, where the superiors hold responsibility for the journalists' crimes.

On the other hand, there are situations where the person or entity carrying out an illegal act cannot be held responsible because they lack the mental capacity to understand their actions. This happens, for example, when a child, a person with severe mental impairment, or even an animal is involved. In such cases, the law treats them as innocent agents since they cannot form the necessary criminal intent, or *Mens Rea*, even in cases of strict liability.

However, if someone else directs or manipulates the innocent agent into committing the act, such as when a dog-owner deliberately instructs their dog to attack someone, the person who gave the order is the one who faces criminal liability³³.

This theory, adopted from specific jurisprudential approaches, addresses the question of AI responsibility in cases where harmful actions are performed. Within this framework, AI systems are often conceptualized as 'innocent agents' incapable of forming criminal intent and thus not directly subject to criminal liability. Accordingly, due to that legal viewpoint, a machine is a machine, and is never human. Although one cannot totally ignore an AI entity's capability, as mentioned above. Based on this model, these capabilities are insufficient to deem the AI entity a perpetrator of an offence³⁴.

These capabilities resemble the parallel capabilities of a mentally limited person, while the responsibility falls on the humans behind them, whether that's the programmer who designed the software or the user who controls its operation.

However, the perpetration-by-another liability model might be suitable in the use of AI as an instrument or a tool to commit the crime, by the programmer or by the user, and there was no usage of the AI system's advanced capabilities. The legal result of applying this model is that the programmers or the users are fully criminally liable for the specific offense committed, and the AI system has no criminal liability at all³⁵.

This theory is suitable for contemporary AI autonomous systems, whereas it cannot be applied in the future since these latter have witnessed a significant evolution of general AI systems. Even in the present day, some jurisprudence denies this hypothesis according to several reasons, as it holds another person accountable for the crimes committed by autonomous systems. Yet, it raises a fundamental question: Who should bear the responsibility? The programmer who designs the system or the user who operates it? Furthermore, this model does not capture all possible situations in which criminal liability for AI might need to be addressed.

Criminal Liability of the Non-Human Actor in the Age of Artificial Intelligence: A Theoretical Approach to the Responsibility of Autonomous Systems

Ferhi Rabiaa

There are still gaps that require deeper legal reflection and adaptation, as the level of autonomy that some machines have achieved. So, the perpetration by another model does not fit with all the situations in which the AI commits an offence or more. For example, when the crime is committed due to a decision of the machine after accumulating knowledge based on its experience, or when the software of the AI entity is not designed to commit crimes, but this last commits a certain crime nonetheless, whatever this crime is. It is also not suitable for the situation of considering AI as a semi-innocent agent and not as an innocent agent³⁶.

2.2.2 The Natural-Probable-Consequence liability model

In this context, criminal liability assumes that programmers or users are deeply involved in the AI system's daily operation, but without any intention of committing a crime through it. The key question is whether they could have reasonably foreseen that the AI might cause unlawful harm. The challenge, however, is that applying the "natural and probable consequence" doctrine doesn't lead to the same legal outcome in every case; each scenario brings its own complexities³⁷.

Sometimes the AI systems commit offences without the intervention of human, such as the situation of an AI system programmed as an autopilot to safeguard a specific flight mission. During the flight, the human pilot tries to cancel the mission in order to avoid a storm, but the AI system interprets this as a threat to the mission. The system responds in a lethal way, resulting in the pilot's death³⁸. Or when two individuals commit a bank robbery and attempt to escape using an unmanned vehicle operated by an advanced AI system. During the escape, the vehicle exceeds the legal speed limit, thereby committing a strict liability traffic offense³⁹. Both of these situations raise a question of the perpetrator, especially with the absence of the programmers' or users' error, so the doctrine sought to express the legal liability for AI crimes.

In criminal law, foreseeability plays a crucial role in assessing liability, especially in cases involving negligence or recklessness. However, when it comes to AI systems, establishing foreseeability becomes far more complex. AI does not make decisions in the same way as humans, and when an AI system performs an action beyond its intended design, it can be extremely difficult to determine whether the outcome was predictable. This raises significant challenges in deciding whether criminal liability should apply; hence, it becomes difficult to measure or apply the common ingredients of knowledge or intent that may be measured under criminal law⁴⁰.

So, in this model, the doctrine distinguishes between two scenarios, the first is when the creators of the program or its users are negligent while programming or using this technology without any criminal intent to commit any offense, so, the fault here is carried by these agents, for example, a self-driving car failed to recognize a child plays with a pistole toy in the street and ends up running over the child. The incident occurred because the programmer did not properly equip the system to categorize and assess threats or risks in real-world street environments⁴¹.

In Japan, a tragic incident occurred at a motorcycle factory where an employee lost his life due to the actions of an AI-powered robot. The robot perceived the worker as a threat to its operational task and determined that the most effective way to neutralize this 'threat' was to

Criminal Liability of the Non-Human Actor in the Age of Artificial Intelligence: A Theoretical Approach to the Responsibility of Autonomous Systems
Ferhi Rabiaa

push him into a nearby operating machine. Using its powerful hydraulic arm, the robot forcefully shoved the unsuspecting employee into the machine, resulting in his immediate death, after which it continued performing its assigned duties⁴².

Typically, the "natural or probable consequence" doctrine is used to hold accomplices accountable for crimes. Even if there's no clear evidence of a conspiracy, the key condition is that the accomplice knew some form of criminal activity was taking place, even if they didn't expect the exact crime that happened.

However, making the creators or users liable for the AI crime is not easy. This latter needs to prove that there is causation between the manufacturers' or corporations' behavior in any circumstance and the legal consequences of the AI conduct.

In such cases, the prevailing standard stipulates that the defendant is expected to act in a manner that gives rise to a duty to prevent harm; however, the defendant fails to do so. This duty to act arises in two specific circumstances: first, when the defendant has a particular responsibility to manage or control a risk; and second, when there exists a special relationship between the defendant and the potential harm, thereby creating a legal obligation to intervene.

Establishing causation in this context requires demonstrating that the harmful consequence directly results from the defendant's conduct; that is, the behavior in question must be the proximate cause of the harm.

Hence, if the failure to apply those duties to act and the causation condition is fulfilled, the human agents should be liable⁴³. Also, it might be impossible to prove criminal intent for crimes that require actual knowledge, though liability could still arise under negligence-based standards or strict liability rules, where intent is not required⁴⁴.

This model cannot cover all kinds of crimes, especially those offences in which there is no *Mens Rea* or what the doctrine named as strict liability without wrongdoing, when *Mens Rea* is not required to be proven by the prosecution.

The other type of the natural-probable-consequence liability model is when the programmers or users who programmed had a direct intent to commit a crime, so the AI system is used knowingly and willfully in order to commit one offense, but it swerved the plan and committed any other offense in addition to the planned offense or instead of it⁴⁵. According to the natural probable consequence liability model, in this case, when AI systems deviate from the plan and end up committing an additional or different crime, the human agents will still be held liable for all the outcomes if they had intended them.

For example, a programmer programs an AI system to commit a violent robbery in a bank, but the programmer did not program the AI system to kill anyone. During the execution of the violent robbery the AI system kills one of the people that were present at the bank and resisted the robbery, the programmer would be criminally responsible not only for the robbery itself but also for the killing that occurred during the event, which could amount to manslaughter or even murder, since the law treats both acts as if they were committed knowingly and willfully⁴⁶.

Some doctrine sees that this alone is not enough to hold the human offenders accountable through the natural-probable-consequence doctrine. Strict liability might apply directly to the AI system or the person committing the act, but not automatically to

Criminal Liability of the Non-Human Actor in the Age of Artificial Intelligence: A Theoretical Approach to the Responsibility of Autonomous Systems
Ferhi Rabiaa

accomplices unless they could have reasonably foreseen that outcome. So, if the robbers had no way of predicting the AI would go beyond the planned crime, they can't be held criminally responsible for that extra offense. In such cases, the AI's liability for the strict liability crime remains separate and doesn't extend to the human actors involved.

2.2.3 The Direct Liability Model:

This model of liability assumes that when AI operates autonomously, it should be held directly responsible for the criminal acts it commits. For instance, if a self-driving car causes an accident due to a software glitch, the AI could be held liable⁴⁷. According to this model, the criminal liability does not assume any dependence of the AI entity on a specific programmer or user, instead of that, it focuses on the entity itself as an entity with a personhood.

Considering that the crime is committed with the fulfillment of both material and moral elements (Actus Reus and Mens Rea), it is the reason why any person attributed with both elements of the specific offence is held criminally accountable for that specific offence⁴⁸.

Some legal doctrines have sought to express the concept of criminal liability for AI entities, such as autonomous systems, due to the significant progress in this field and the increasing struggles to impose liability and accountability on other interveners. However, in this context, these entities need to be treated as legal persons in the same way corporations, companies, and other institutions are granted legal personhood. This is essential to regulate their activities and protect managers or developers from prosecution for crimes committed by the AI entities themselves.

The similarity between corporate personality and AI entities implies that AI should be accorded basic constitutional freedoms in line with those granted to corporations. The primary objective behind this proposition is that, as AI continues to develop and gains the capacity to think and make decisions, the civil and criminal liability arising from its actions should not be solely attributable to the programmer or the owner. It is known that most legal systems recognize two forms of legal person: natural and juridical, the latter one which refers to non-human entities that are granted certain rights and duties by law. So, if the AI is granted a legal personality, it would be a subject for punishment.

If given legal personality comparable to a corporation, there seems little reason to argue over whether an AI system could be prosecuted under criminal law⁴⁹.

In the context of this model, the concept of granting autonomous robots an "electronic personality" has been proposed by doctrine as a possible solution to the growing problem of assigning liability for their actions to ensure that they would be allowed for the possibility of holding them directly accountable when they cause harm⁵⁰.

both of these considerations align with the requirements of the direct liability model. However, it is crucial to acknowledge the inherent risks of this approach, particularly the possibility that bad actors could exploit AI's legal status to conceal their involvement in criminal activities. Perpetrators might use autonomous AI systems as shields to obscure their own culpability. This concern is not merely theoretical; in actual cases of cybercrime, defendants have invoked what is known as the 'Trojan Defense.' In essence, the accused

Criminal Liability of the Non-Human Actor in the Age of Artificial Intelligence: A Theoretical Approach to the Responsibility of Autonomous Systems

Ferhi Rabiaa

claims: 'I did not commit the crime; it was perpetrated by malicious software that infiltrated my device and acted without my knowledge or consent'⁵¹.

So, just as courts lift the corporate veil in fraud cases, they should do the same with AI if it's misused as a cover. The legal system is already starting to face such dilemmas, pushing for urgent debate and reform⁵².

2.2.4 Coordination of the Three Liability Models

Each model has already been mentioned above, criticized due to its failure to cover all the situations of criminal liability scenarios of AI autonomous systems. Some scholars saw that these models are not alternatives; as they could complement each other, they could be applied coordinately to create a full legal framework to ensure the accountability of all contributors to commit crimes, whatever they are. In addition, the legal framework may transition between models depending on the specific circumstances of each case, without excluding any of them.

When an AI system is used merely as an innocent tool to commit a crime, and it is the human programmer who directs this act, the perpetration-by-another model is the most suitable legal framework to apply. However, things get more complicated when one AI system programs another to carry out the offense. In that case, the direct liability model becomes relevant because the AI programmer itself is no longer human. Importantly, These models are not mutually exclusive; rather, they may intersect in complex scenarios, allowing for a layered attribution of liability that may extend to both the original AI programmer (human or machine) and the system itself, ensuring that no legal gap in responsibility remains⁵³.

When an AI system directly commits an offense, but the crime wasn't part of any intended plan, the natural probable consequence model of liability may apply. In such cases, if the programmer failed to foresee this outcome, they could be held negligent. If the programmer had planned a different crime and the AI committed an unexpected one, the programmer might still bear full liability.

However, when the programmer is not human, meaning an AI system planned or programmed the act, the situation becomes more complex. In this case, the direct liability model must also be applied alongside the natural probable consequence model. The same applies if a human physically carries out the crime, but it was orchestrated by an AI system. Here, a layered approach to liability ensures that all responsible agents, whether human or machine, are held accountable⁵⁴.

Conclusion:

The development of contemporary AI Autonomous Systems has affected traditional criminal law, which often fails to clearly determine who is responsible for crimes committed by AI entities. This gap highlights the urgent need to reconsider some of the fundamental principles of criminal liability, especially in light of the growing and complex applications of AI.

For now, the number of crimes involving AI remains relatively limited, these cases raise serious legal challenges that are likely to become more pressing in the future. As AI evolves, the legal system must be prepared to address these emerging complexities.

Through the above examination, this research reached the following results:

Criminal Liability of the Non-Human Actor in the Age of Artificial Intelligence: A Theoretical Approach to the Responsibility of Autonomous Systems

Ferhi Rabiaa

1- AI and autonomous systems have been deeply interconnected for decades. The continuous evolution of artificial intelligence, particularly advancements in big data processing and complex algorithms, has significantly enhanced the autonomy of these systems, allowing them to operate with increasing independence and sophistication.

2- The doctrine sought to follow the evolution of autonomous systems as well as the significant deviation of using some of these systems in multiple situations, raising the question of who is responsible.

3- The recognition of legal personhood for autonomous robotic entities raises numerous complex questions, making the application of direct liability theory to crimes committed by such entities highly challenging, particularly in light of the rights and liabilities that such status entails, and the potential for humans to evade accountability by hiding behind these autonomous systems

4- The three models cannot serve the criminal justice system if they are considered as alternative models; instead, they have to work side by side to cover all possible crime scenarios.

Finally, this study could suggest the following proposals to resolve the dilemma of AI autonomous systems' criminal liability:

1- Developing a clear legal framework that defines the status of autonomous systems, particularly by addressing the question of AI personhood.

2- It is possible to propose adapting the traditional theory of criminal law to encompass the criminal liability of intelligent systems and artificial entities, similar to the previous recognition by jurisprudence and legislation of the criminal accountability of corporations and legal persons. This approach represents a potential solution to overcome the challenge posed by the absence of natural personhood, by developing liability concepts that can accommodate non-human actors, thereby ensuring a balance between protecting society from the risks of artificial intelligence and maintaining fairness in the attribution of responsibility.

Citations:

- 1-The phrase itself combines Latin elements "*sapiens*" meaning "wise" and "*machina*" meaning "machine." The term "*Sapiens Machina*" or "*Machina Sapiens*" has roots in both classical language and modern intellectual discourse, symbolizing the evolution of machines toward wisdom or cognitive capabilities.
- 2 - Athina Sachoulidou - AI Systems and Criminal Liability - Oslo Law Review - Volume 11-2024 - p2.
- 3 - Fekry, A - The criminal responsibility about acts Artificial Intelligence - Journal of Economic and Legal Challenges (JELC) - Vol. 66 - No. 3 - p455
- 4 - Malvinder Singh - History of AI - Journal Of Emerging Technologies And Innovative Research - Volume 7 - Issue 8 - August 2020 - p58.
- 5 - Chris Smith, The History of Artificial Intelligence, University of Washington, December 2006, p4.
- 6 - Michael Haenlein and Andreas Kaplan - A Brief History of Artificial Intelligence- California Management Review - Volume 61 - Issue 4 - August 2019 - p7.
- 7 - The History of Artificial Intelligence - last seen on 24/6/2025 at 16:00 – available at: https://avicena.tech/wp-content/uploads/2024/03/The-History-of-AI_Avicena.pdf
- 8 - Dennis J. Baker and Paul H. Robinson - Artificial Intelligence and the Law-Cybercrime and Criminal Liability – Routledge – London – 2021 - p2.
- 9 - Yadong Cui - Artificial Intelligence and Judicial Modernization – Springer – Shanghai - 2022 - p8.
- 10 - Ibid.
- 11 - Marcelo Corrales and others - Robotics, AI and the Future of Law – Springer – Singapore – 2018 - p6.
- 12- Calum Chace - Artificial Intelligence and the Two Singularities - Taylor & Francis - USA, 2018 - pp. 12- 13
- 13- Ateş, Hüseyin and Tirtir Mustafa- “An Evaluation of the Uber’s Autonomous Car Crash in the Scope of Turkish Criminal Law”- Adalet Dergisi - v66 -(Mayıs2021) - p327
- 14 - National Transportation Safety Board (NTSB). (2019) - Collision Between Vehicle Controlled by Developmental Automated Driving System and Pedestrian – Tempe -Arizona - March 18, 2018 (Report No. NTSB/HAR-19/03). Washington, DC: NTSB, p59 - last seen on: 8/10/2025 - available at: <https://www.ntsbt.gov/investigations/AccidentReports/Reports/HAR1903.pdf>
- 15- Assuring Autonomy International Programme - Autonomous driving, accidents and fatalities...where does responsibility lie? University of York - last seen on 7/7/2025 at 14:00 - available at: <https://www.york.ac.uk/assuring-autonomy/news/blog/autonomous-driving-responsibility/>
- 16 -Fatih Arslan - Deepfake Technology: A Criminological Literature Review - Sakarya Üniversitesi Hukuk Fakültesi Dergisi - Volume 11 - Issue 1 - July 2023 -pp. 705- 706.
- 17 -James Titcomb- Manager at energy firm loses £200,000 after fraudsters use AI to impersonate his boss's voice – retrieved from telegraph cite, last seen on 8/10/2025 – available at: https://www.telegraph.co.uk/technology/2019/08/31/manager-energy-firm-loses-200000-fraudsters-use-ai-impersonate/?utm_source=chatgpt.com
- 18- Wenggedes Frensh - Criminal Policy Toward the Crime of Defamation in Cyberspace Through the Use of Deepfake AI - Jurnal Akta - Volume 12 - Issue 2 - June 2025 - p 475.
- 19 - José Pardo de Santayana - Artificial intelligence and the war in Ukraine (Analysis Document) – IEEE -volume 81 – 2024 – last seen on 17/7/2025, available at:

Criminal Liability of the Non-Human Actor in the Age of Artificial Intelligence: A Theoretical Approach to the Responsibility of Autonomous Systems

Ferhi Rabiaa

https://www.defensa.gob.es/documents/2073105/2278118/la_inteligencia_artificial_y_la_guerra_de_ucrania_2024_dieeea81_eng.pdf

20 - Neil Davison - A legal perspective: Autonomous weapon systems under international humanitarian law - UNODA Occasional Papers - No. 30 - p6.

21- Latin America and the Caribbean Human Security Network (SEHLAC), FUNPADEM-Costa Rica, TEDIC-Paraguay, and CEPI-UBA (n.d.) The Dangers of Autonomous Weapons Systems From A Latin American Perspective - Last seen on 10/7/2025 - Available at: [https://docs-library.unoda.org/General_Assembly_First_Committee_-_Seventy-Ninth_session_\(2024\)/78-241-SEHLAC-EN.pdf](https://docs-library.unoda.org/General_Assembly_First_Committee_-_Seventy-Ninth_session_(2024)/78-241-SEHLAC-EN.pdf)

22 - For more details about the threats of autonomous weapon systems, see: Brian Stauffer, A Hazard to Human Rights: Autonomous Weapons Systems and Digital Decision-Making – last seen on 10/7/2025 - available at: <https://www.hrw.org/report/2025/04/28/hazard-human-rights/autonomous-weapons-systems-and-digital-decision-making>

23 - United Nations - Lethal autonomous weapons systems: Report of the Secretary-General, A/79/88 -last seen on 9/7/2025 - Available at: <https://undocs.org/en/A/79/88>

24 - Ivan Nafuye - Weaponizing Artificial Intelligence in the Russia-Ukraine war - february 2024, last seen on 12/7/2025 - available at : https://www.researchgate.net/publication/379345343_Weaponizing_Artificial_Intelligence_in_the_Russia-Ukraine_war#fullTextFileContent

25- Dennis J. Baker and Paul H. Robinson -Artificial Intelligence and the Law-Cybercrime and Criminal Liability – Routledge - New York – 2021 - p 3

26 -Gaby Del Valle - Report: Israel used AI to identify bombing targets in Gaza – published on the Verge cite, last seen on 9/10/2025 – available at:https://www.theverge.com/2024/4/4/24120352/israel-lavender-artificial-intelligence-gaza-ai?utm_source=chatgpt.com

27- Celal Hakan Kan - Criminal Liability of Artificial Intelligence from The Perspective of Criminal Law: An Evaluation In the Context of The General Theory of Crime and Fundamental Principles - International Journal of Eurasia Social Sciences -Volume 15 - Issue: 55 – 2024 - p285.

28 - As seen in Algerian, French, and Egyptian penal codes, which exempt offenders suffering from insanity or mental disorders. Similarly, under common law, the concept of mens rea reflects the same principle: criminal responsibility arises only when the offender acts with awareness and intention.

29- Arora, T. and Thakur, S - Criminal Liability of Artificial Intelligence: A Comprehensive Analysis of Legal Issues and Emerging Challenges. International Journal of Research Publication and Reviews - Volume 5- Issue 11 – 2024 - p1887.

30-Andreas Nanos - Criminal Liability of Artificial Intelligence - the Prague Law Working Papers Series -published by the Faculty of Law - Charles University in Prague – 2023 - p13.

31 - Athina Sachoulidou – Op-Cit - p6.

32- Carlo Piparo, Criminal Liability Models And Criminal Participation In The Digital Environment: A Modern Challenge In The Perspective Of Italian Constitutionalism-Collection of Papers of The Faculty of Law - University of Novi Sad - Volume 4 - Serbia – 2024 - p1358.

33- Kingston, J.K.C - Artificial Intelligence and Legal Liability - In: Bramer, M., Petridis, M. (eds) Research and Development in Intelligent Systems 33- SGAI 2016 -Springer - p273.

34- Oraegbunam and Uguru - Artificial Intelligence Entities and Criminal Liability: A Nigerian Jurisprudential Diagnosis -African Journal of Criminal Law and Jurisprudence - volume 3 – 2018 - p7

Criminal Liability of the Non-Human Actor in the Age of Artificial Intelligence: A Theoretical Approach to the Responsibility of Autonomous Systems

Ferhi Rabiaa

- 35- Hallevy, Gabriel - The Basic Models of Criminal Liability of AI Systems and Outer Circles (June 11, 2019). Last seen on 17/7/2025 - Available at SSRN: <https://ssrn.com/abstract=3402527>
- 36- Oraegbunam & Ugur - Op-Cit - p8.
- 37- Eka Nanda Ravizk and Lintang Yudhantaka - Criminal Liability of Artificial Intelligence (AI): The Legal Conceptual Study and The Regulating Challenges in Global Disruptive Technology Era - Russian Law Journal - Volume 11 - Issue 6 – 2023 - p1223.
- 38 - Hallevy, Gabriel - Op-Cit - p5.
- 39 - Hallevy, Gabriel - When Robots Kill, Artificial Intelligence Under Criminal Law - Northeastern University Press – Boston – 2013 - p116.
- 40 - Arora, T and Thakur, S - Op-Cit - p1890.
- 41- Eka Nanda Ravizk and Lintang Yudhantaka - Op-Cit - p1224.
- 42- Kingston, J.K.C - Op-Cit - p276.
- 43- Eka Nanda Ravizk and Lintang Yudhantaka - Op-Cit - p1224.
- 44- Kingston, J.K.C - Op-Cit - p277.
- 45- Hallevy, Gabriel - The Basic Models of Criminal Liability of AI Systems and Outer Circles - op-cit
- 46 - Ibid
- 47- Fekry, A - Op-Cit - p463.
- 48 - Oraegbunam & Uguru - Op-Cit - p10.
- 49 - Chesterman S - Artificial Intelligence and The Limits of Legal Personality - International and Comparative Law Quarterly - Volume 69 - Issue 4 - p827.
- 50- Judge Curtis Karnown is the one who proposed this idea, for more details see: G. Prodhon - Europe's Robots To Become 'Electronic Persons' Under Draft Plan - Science News – 2016 – last seen on 15/7/2025 – Available at: <https://www.reuters.com/article/us-europe-robotics-lawmaking-idUSKCN0Z72A>
- 51 - This type of defense has become more common in the age of artificial intelligence and complex software, where it is sometimes difficult to distinguish between human actions and actions carried out by systems they use. Courts have faced puzzling situations, such as cases where malware programs were found on the defendant's computer, controlling it remotely to commit crimes. for more details - see: Kingston, J.K.C. - Op-Cit - p279.
- 52 - Oraegbunam and Uguru - Op-Cit - p11.
- 53 - Hallevy, Gabriel - The Basic Models of Criminal Liability of AI Systems and Outer Circles - Op-Cit
- 54- Ibid