DISCUSSION PAPER

From Code to Courtroom:
Legal Accountability in Artificial Intelligence

İhsan Faruk Kılavuz

TRT WORLD research centre

DISCUSSION PAPER

From Code to Courtroom:
Legal Accountability in Artificial Intelligence

İhsan Faruk Kılavuz

TRT WORLD research centre

WRITTEN BY

İhsan Faruk Kılavuz

PUBLISHER

TRT TRAINING AND RESEARCH DEPARTMENT

May 2024

TRT TRAINING AND RESEARCH DEPARTMENT

AHMET ADNAN SAYGUN STREET NO:83 34347

ULUS, BEŞİKTAŞ

İSTANBUL / TÜRKİYE

TRT WORLD LONDON

PORTLAND HOUSE

4 GREAT PORTLAND STREET NO:4

LONDON / UNITED KINGDOM

TRT WORLD WASHINGTON D.C.

1819 L STREET NW SUITE 700 20036

WASHINGTON DC

researchcentre.trtworld.com

Introduction

The digital age has been significantly transformed by advancements in cognitive science, leading to the emergence of artificial intelligence (AI) as a pivotal force in modern society. AI systems, characterised by their ability to learn and adapt through sophisticated algorithms and machine learning techniques, have rapidly evolved to perform tasks with human-like proficiency. These systems have become integral across diverse sectors, highlighting their broad impact on everyday life. The profound integration of AI into daily areas underscores the need to examine its societal implications comprehensively.

As AI continues to develop autonomous capabilities, it introduces a range of benefits and risks. The ability of AI systems to make independent decisions raises critical concerns about accountability, especially when these decisions result in legal or ethical violations. Incidents involving AI, such as autonomous vehicles causing accidents or biased algorithms leading to discriminatory practices, illustrate the pressing need for legal frameworks that can address these challenges.

This paper aims to explore the necessity for robust legal structures to govern AI technologies, ensuring that their deployment maximizes societal benefits while mitigating potential harms. By investigating current legal approaches and proposing innovative solutions, this discussion seeks to contribute to the development of a coherent and effective regulatory environment for AI.

(Tayfun Coşkun - Anadolu Agency)

Artificial Intelligence and the Need for Legal Frameworks

Advancements in cognitive science have ushered in the era of artificial intelligence in the contemporary digital world. AI entities have demonstrated a significant ability to learn from their experiences through various algorithmic structures and sophisticated software, particularly leveraging machine learning methodologies.

Over time, they have developed autonomous and cognitive capabilities, exhibiting human-like skills. Consequently, AI-based systems have seamlessly integrated into nearly every aspect of human life. From healthcare, education, and production to agriculture, finance, automotive, and defence sectors, it is challenging to find a domain untouched by artificial intelligence. (Sadi, 2022) In other words, artificial intelligence and AI-based systems have transcended the realm of science fiction to become indispensable realities of our lives. Therefore, evaluating the impact of AI systems necessitates considering not only from an individual perspective but also from the broader perspective of society.

On the other hand, the increasing integration of artificial intelligence into human life and its resulting impact on societal dynamics have necessitated their legal scrutiny because the complexity of AI systems endowed with the ability to make autonomous decisions brings various risks alongside their benefits. Put differently, AI entities' autonomous decisions and actions sometimes violate individuals' property and personal values, thus leading to legal infringements. (Report, 2020)

The proliferation of autonomous systems has brought various legal and ethical challenges concerning the damages and legal violations they may engender. One notable example involves a case where an autonomous vehicle was involved in a fatal accident. Despite its advanced sensors and algorithms, the vehicle failed to accurately detect a pedestrian crossing the street, leading to a tragic outcome. In this instance, questions arose regarding the liability for the accident. Should it fall on the vehicle manufacturer, the AI system developer, or the human operator who was meant to supervise the vehicle's operation?

Furthermore, concerns have been raised about using AI in decision-making processes, particularly in areas such as finance and hiring. There have been instances where AI algorithms, trained on biased data sets, have perpetuated discriminatory practices, leading to legal challenges and debates about accountability. For example, in a high-profile case, a financial institution was accused of using AI algorithms that systematically denied mortgage applications from certain demographic groups, resulting in allegations of unlawful discrimination.

Moreover, the increasing reliance on AI in fields like healthcare has also raised questions about the potential for medical malpractice. In one instance, an AI-powered diagnostic tool misdiagnosed a patient's condition, leading to delays in treatment and exacerbating their health issues. This sparked a legal battle over accountability for the erroneous diagnosis, highlighting the complexities of attributing responsibility in cases involving AI systems. (Kılıçarslan, 2019)

These examples underscore the urgent need for robust legal frameworks to govern AI technologies' development, deployment, and use. (Kılıçarslan, 2019)

Determining Blame: AI Systems or Human Operators?

Given the absence of specific legislation holding AI directly accountable, compensation for the harm resulting from AI failure typically relies on the conventional "negligent tort liability" framework.

Under certain conditions, parties such as developers, manufacturers, and users may be held responsible for AI-induced damages within this framework.

These conditions for attributing liability to these "other" parties include:

- The defendant has a legal obligation to exercise care.
- The defendant's failure to fulfil this duty.
- The resulting harm suffered by the plaintiff due to this breach.

Consequently, the plaintiff must demonstrate that the defendant—i.e., the other party—owed a duty of care regarding the AI system, violated that duty, and that this violation directly led to the plaintiff's damage. Ultimately, liability for negligence would lie with the person, persons or entities who caused the damage or who might have foreseen the product being used in the way that it was used.

However, the evolution of autonomous functionalities within AI systems empowers them to autonomously make decisions and act beyond the scope anticipated by their developers, manufacturers, and users. This evolution renders the conventional liability framework insufficient when applied to modern AI systems.

To put it differently, if the damage stems from AI behaviours entirely unforeseen, negligence claims may encounter challenges due to the lack of foreseeability, potentially resulting in no party being held liable. Therefore, establishing proximity and foreseeability becomes more challenging in cases where AI systems operate autonomously or diverge significantly from human decision-making processes.

This scenario raises doubts about artificial intelligence's accountability. The central debate on AI accountability revolves around the concept of "its legal personhood." (Kılıçarslan, 2019)

Is it Necessary to Legally Grant Personhood to AI?

The effective integration of AI systems and robotic technology into social and economic life has rendered social relationships and resulting legal relationships too intricate and complex to be managed solely by natural and legal persons. (Dogan, 2002) Additionally, anticipated changes in interaction patterns between humans and AI entities, along with expectations and demands regarding the functional domains of AI, underscore the importance of determining the legal status of these entities. Therefore, progressing towards new and comprehensive codifications encompassing not only humans and current legal entities but also artificial intelligence appears inevitable within traditional legal frameworks. Indeed, the European Parliament's call to the European Commission to prepare legislation addressing the challenges AI will face in the future further validates this need.

Modern societies are constantly in flux because of fast-paced scientific and technological advancements and legal frameworks often lag behind. Until regulations catch up, ethics must guide the community. This statement is also valid for the development of AI. (Henz, 2021)

However, relying solely on "ethical values" for AI is no longer sufficient. With the progress demonstrated by AI and the increasing complexity of systems, legal issues transcending ethical concerns have begun to arise. Consequently, discussions concerning artificial intelligence should move beyond the dimension of "ethics" and extend into the "legal regulatory frameworks" realm.

In essence, the issue of holding artificial intelligence legally accountable fundamentally entails granting this entity a set of rights and corresponding obligations, leading to various perspectives on whether legal personhood should be recognised for AI.

Views Rejecting the Granting of Legal Personhood to AI

Various views exist to reject the attribution of legal personhood to artificial intelligence entities.

One such perspective is the moral personality view, which asserts that only humans possess consciousness, the ability to think and plan, innate intelligence, emotions, and physical capacities, thus constituting personality as a reflection of inherent intelligence and internal skills. Even though artificial intelligence may mimic human intelligence, consciousness, and emotions, it cannot truly possess intelligence and emotions in the real sense. Therefore, personality applies only to humans. (Dogan, 2022)

It is worth noting that this approach needs to be revised to explain the legal existence of entities such as associations and corporations, as it merely reduces the concept of personality to humans. Thus, basing legal principles on this approach seems implausible.

Another basis for rejecting the attribution of legal personality to artificial intelligence is the argument that such status is unnecessary. According to this perspective, legal issues arising from artificial intelligence's autonomy can be addressed through certain insurance systems without the need to confer legal personality upon it. Therefore, there is no need to elevate AI entities beyond the status of "property." (Sadi, 2022) In other words, this view acknowledges the autonomy of artificial intelligence while suggesting that they may only be objects rather than subjects of rights or responsibilities. (Kılıçarslan, 2019)

The final view opposing the recognition of legal personality posits that such recognition would conflict with human interests, particularly in areas such as employment and security. However, it is important to note that this perspective addresses the issue on a political rather than a legal basis.

Overall, approaches rejecting the attribution of legal personality to artificial intelligence view personality as a set of innate values acquired at birth. Considering personality as an inseparable element of fundamental rights and duties, this perspective acknowledges that artificial intelligence cannot fulfil these rights and responsibilities.

Views Attributing Legal Personhood to AI

The concept of granting legal personhood to artificial entities is rooted in the idea of recognising rights and corresponding obligations within the framework of legal policies set by lawmakers. Approaches advocating for granting legal personhood to artificial intelligence argue that if it gains societal acceptance and is in line with general legal guidelines, it can be recognised as a legal actor.

Granting legal personhood entails acknowledging their rights and, consequently, their responsibilities under positive law. (Doğan, 2022) Indeed, the orthodox approach emphasises that being entitled to rights and responsibilities under the law is inherently intertwined with possessing a legal personality. (Kurki, 2019)

Certain criteria are sought for attributing personhood to artificial intelligence entities: the ability to engage in environmental interactions, complex thinking and communication, the capacity to implement decisions, and a level of consciousness to sustain their presence in society.

The primary motivation behind attributing personhood is the necessity of having a party to compensate for damages in case increasingly complex AI entities cause harm. Furthermore, another motivation would contribute to economic development in countries where autonomous systems are utilised. For instance, within the United States, 47 states have ratified "the Uniform Electronic Transactions Act (UETA)", legislation initially introduced by the District of Columbia and the Virgin Islands. This regulation permits contracts generated by AI-enabled devices to represent a tangible legal advancement in ascribing "limited legal personality" to such entities.

The views accepting the personhood of artificial intelligence also vary within themselves.

1) Slavery Approach

The initial perspective advocating for the attribution of personhood to AI draws parallels with the legal status of enslaved people n Roman law, referred to as the "slavery approach". This approach posits that AI devices occupy a societal position akin to that of enslaved people in ancient times. (Ozgun Law)

In Roman legal tradition, enslaved people were regarded as a form of "property" possessing cognitive abilities and decision-making capacity. Similarly, this viewpoint classifies AI devices as modern manifestations with technological capabilities resembling those of ancient slaves. Consequently, the legal treatment afforded to enslaved people could be extended to AI entities.

Under this framework, rights and obligations concerning AI systems would be assigned to their "owners". In essence, akin to the responsibilities associated with slave ownership, users of AI would assume liability for any resultant harm caused by these systems.

While this perspective offers a structured basis for holding AI accountable, its widespread acceptance is hindered by the lack of relevance of slavery and its implications in contemporary legal frameworks. Such frameworks, which diverge from universal legal principles, undermine the universal applicability of this approach.

2) Legal Personality

In the discourse surrounding the legal accountability of artificial intelligence, there is a proposition suggesting that AI entities could be endowed with legal personhood akin to corporate entities rather than being treated merely as inanimate objects or machines. While AI lacks the innate characteristics of natural persons, such as human consciousness, its distinct ability to make autonomous decisions sets it apart from traditional objects.

This proposal draws parallels with legal systems that extend personhood to entities beyond individuals, such as corporations, partnerships, and other organised groups. By granting legal personality to AI entities, proponents argue for establishing a legal framework that treats them as distinct entities with rights and responsibilities. In this proposed framework, the relationship between AI and its creator or user resembles that of a company and its board of directors. Consequently, assigning legal personhood to AI would

entail holding its producer or user accountable for its actions, akin to the oversight responsibilities of a board of directors for a corporation's activities. (Sadi, 2022)

3) The non-human Personality Proposal

Another perspective regarding the legal classification of artificial intelligence is the "non-human person" proposition. This approach acknowledges the cognitive capabilities of AI systems, albeit distinct from those of humans, and proposes granting them a legal status separate from "natural persons."

Advocates of the "non-human person" approach propose endowing AI with a *sui generis* personality. However, this stance has faced criticism due to the vagueness and conceptual nature of this *sui generis* personality and its failure to offer a precise framework for holding AI accountable within legal contexts.

4) The European Parliament Report and Electronic Personality Proposal

The discourse surrounding the legal personality of artificial intelligence is a significant subject of consideration not only on a global scale but also within the European context, where governmental bodies and other entities are actively involved in various research endeavours and initiatives. The European Parliament report on January 27, 2017, reflects a notable effort.

This report introduces several suggestions and viewpoints concerning the attribution of personhood to AI, with one particularly noteworthy proposal being the proposition for a novel legal classification termed "electronic personality," distinct from the existing categories of natural and legal personalities. Recognising the current inability of AI to possess the legal status accorded to humans, the report suggests exploring alternatives such as the electronic personality concept. This entails the registration of each AI entity in an official registry and the establishment of dedicated funds related to them. Consequently, in instances where AI causes harm, compensation would be facilitated through these funds, operating under a principle of strict liability wherein establishing causation between the damage and the actions of the AI would suffice to establish liability. (Kılıçarslan, 2019)

The approach delineated in the report allows for a gradual transition of liability onto AI entities as their learning capabilities and autonomy expand. Consequently, as AI autonomy in decision-making advances, the accountability for the outcomes of these decisions cannot be solely attributed to producers, developers, and users.

Compared to other propositions, this approach distinguishes itself with its tangible solution of registering each AI entity and facilitating compensation through dedicated funds, thereby firmly anchoring the principle of strict liability within a legal framework.

AI-Powered Robot named "Marbot" is seen at the entrance of pharmacy as he helps customers like a team member in Yalova, Turkiye on August 29, 2023. The robot known as 'Marbot,' equipped with an artificial intelligence system and stationed at the entrance of a Yalova pharmacy, is capturing the interest of customers. Robot possesses the capability to identify inquiries from customers as they enter the pharmacy and respond to them using the information provided by the company. (Sıtkı Yıldız - Anadolu Agency)

The Other Side of the Coin: The Criminal Liability Dimension

As explored in the preceding section, civil law liability delves into the realm of compensatory measures, aiming to restore or compensate for damages caused by AI entities' actions. This framework primarily operates within the context of tort law, where individuals or entities are held accountable for their actions, often resulting in financial restitution to the aggrieved parties.

On the other hand, the transition from civil to criminal liability introduces additional complexities. While civil liability may suffice for addressing individual grievances and providing restitution, criminal liability serves a broader societal function by promoting accountability, deterrence, and the maintenance of public order. However, applying criminal law to AI systems raises intricate questions regarding intent, culpability, and the attribution of responsibility in cases where autonomous entities operate beyond human control or oversight.

Criminal laws generally require both an *actus reus* (an action) and a *mens rea* (a mental intent). The legal controversy regarding artificial intelligence's criminal liability is in the mens rea element rather than the actus reus. At this juncture, various propositions have been put forth to address artificial intelligence's accountability for actions that constitute crimes and lead to criminal liability, distinct from civil liability.

1)Perpetrator-via-another

The first approach is the "innocent agent" approach, which is based on a legal doctrine applicable to offences committed by individuals with mental impairments, children, or animals. In cases where a crime is committed by someone with cognitive deficiencies, for instance, a child, the perpetrator is considered an innocent agent because this person lacks the mental capacity to form *mens rea*. Therefore, individuals in these categories cannot be criminally liable for their actions.

However, if the innocent agent is instructed by another individual (for instance, if a dog owner instructs their dog to attack someone), the person giving the instructions bears criminal responsibility.

According to this model, AI systems could be regarded as innocent agents, with either the software programmer or the user deemed the "perpetrator-via-another." Consequently, under this perspective, other individuals, such as developers and users, could be subject to criminal liability for the criminal acts committed by artificial intelligence.

2)Natural-probably-consequence

In the natural-probable-consequence model, a segment of an AI program, initially designed for benign purposes, is activated inappropriately, and ends up committing a criminal action. A well-known example of this occurred when a Japanese worker in a motorcycle factory was fatally injured by an artificially intelligent robot operating nearby. The robot mistakenly perceived the worker as a threat to its mission and decided to push him into an adjacent operating machine to neutralise this perceived threat. Utilising its powerful hydraulic arm, the robot forcefully pushed the worker into the machine, resulting in his immediate death, after which it resumed its regular duties. (Kurki, 2019)

Under normal legal circumstances, the concept of "natural or probable consequence" liability is typically applied to prosecute accomplices to a crime. Even in cases where no formal conspiracy can be proven, an accomplice may still be held legally liable if the criminal actions of the perpetrator were a foreseeable outcome of a scheme that the accomplice either encouraged or assisted in, provided that the accomplice was aware of the ongoing criminal activity. Therefore, users or, more likely, programmers could potentially be held legally accountable if they were aware that a criminal offence was a foreseeable consequence of their programs or the use of an application. (Kingston, 2018)

However, it is crucial to differentiate between AI programs specifically programmed to engage in criminal activities and those designed for entirely different purposes. Crimes requiring a specific *mens rea*, such as knowledge, may not be prosecutable for

the latter group of programs. Nonetheless, offences involving a "reasonable person" *mens rea* or strict liability offences may still be applicable.

3)Direct Liability

In this model, both *actus reus* and *mens rea* are attributed to an AI system.

Determining *actus reus* in an AI system is relatively straightforward. If a system undertakes an action leading to a criminal act or fails to act when obligated to do so, the *actus reus* of the offence is present.

However, assigning *mens rea* is more complex, and this is where the three levels of *mens rea* become significant. For strict liability offences, which do not require intent to commit the offence, holding AI programs criminally liable may be feasible. For instance, in the case of self-driving cars, speeding constitutes a strict liability offence. Consequently, according to a view, if a self-driving car is found to exceed the speed limit, the law may assign criminal liability to the AI program controlling the vehicle at that time. (Kingston, 2018)

This possibility raises several other considerations. These include potential defences: Could a malfunctioning program claim a defence similar to the human defence of insanity? Or, if affected by an electronic virus, could it assert defences akin to coercion or intoxication? Additionally, questions regarding punishment emerge: In cases where an AI system is directly liable for an offence, who or what would be subject to punishment? How will the punishment be executed? Will the plug of the artificial intelligence device be pulled?

Therefore, the practical implementation of this approach, which directly holds artificial intelligence responsible, is not feasible.

Conclusion

In conclusion, the rapid integration of artificial intelligence into various sectors of society presents both opportunities and challenges. The autonomous capabilities of AI systems necessitate a comprehensive legal framework to address the ethical, legal, and societal implications of their deployment. Current negligence-based liability frameworks are insufficient to handle the complexities arising from AI's autonomous decision-making. This paper has explored various perspectives on AI accountability, ranging from the necessity of granting AI legal personhood to alternative liability models that ensure responsible usage and compensation for harm. As AI continues to evolve, it is imperative for lawmakers, technologists, and society to develop legal structures that balance innovation with accountability collaboratively. Recognising AI's potential to transform society, we must proactively establish legal mechanisms to mitigate risks, protect individual rights, and promote ethical standards. By doing so, we can harness the benefits of AI while safeguarding against its potential harms, ensuring a future where technology serves humanity's best interests within a robust legal framework.

Bibliography

Dogan E, (2022), Search for a new legal personality in digital age: artificial intelligence, *Journal of the Turkish Bar Association* https://search.trdizin.gov.tr/tr/yayin/detay/1134981/dijital-cagda-yeni-bir-hukuki-kisilik-arayisi-yapay-zeka

Henz P, (2021, September) Ethical and legal responsibility for artificial intelligence, *Discover Artificial Intelligence*, https://link.springer.com/article/10.1007/s44163-021-00002-4

Kılıçarslan Kara S., (2019), Legal status of artificial intelligence and debates on its legal personality, *Yildirim Beyazıt Law Journal*, https://dergipark.org.tr/tr/pub/ybuhukuk/issue/47456/599224

Kingston J.K.C, (2016, November), Artificial intelligence and legal liability, *International Conference on Innovative Techniques and Applications of Artificial Intelligence*, https://www.researchgate.net/publication/309695295_Artificial_Intelligence_and_Legal_Liability

Kurki Visa A.J, (2019, August), The legal personhood of artificial intelligences, *Oxford Academic*, https://academic.oup.com/book/35026/chapter/298856312

Report on the Safety and Liability of AI, the Internet of Things and Robotics, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0064

Sadi S, (2022, February), Legal personality of artificial intelligence systems, *Okan University Law Bulletin*, *https://okan.edu.tr/hukuk/sayfa/8088/yapay-zeka-sistemlerinin-hukuki-kisiligi/*

Ozgun Law, An evaluation on the opinions on the legal status of artificial intelligence, https://www.ozgunlaw.com/makaleler/yapay-zekanin-hukuki-statusu-hakkindaki-gorusler-uzerine-degerlendirme-744

TRT WORLD
research centre

TRTWORLD
research
centre