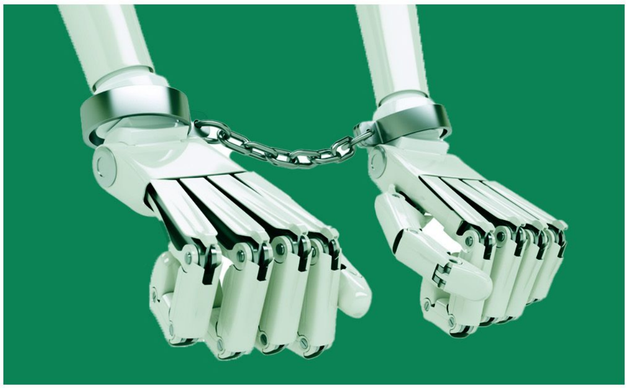


Expert Policy Research Article

DOC ILA - NEW TECH/AI/06

Can Artificial Intelligence be Liable for a Crime?

Francis Aondongu Wayo

Introduction

In the opening preface of his seminal work "When Robots Kill: Artificial Intelligence Under Criminal Law", Gabriel Hallevy recounts a 1981 story in which an artificial intelligence (AI) robot killed a Japanese employee in a motorcycle factory having erroneously identified the employee as a threat to its mission.[1] In such situation, Hallevy further interrogates who will "be held criminally liable for this homicide".[2] This is a complicated question that finding adequate answers through the extant criminal law jurisprudence may seem too daunting a task or a Sisyphean one. However, as AI is increasingly becoming an integral part of society, there is need to understand the place of law in relation to the conduct of AI. Consequently, this paper delves into the interface between AI and criminal jurisprudence with particular focus on examining the culpability of AI under criminal law. In doing this, the paper, inter alia addresses the following questions.

a. What is AI?
b. Can AI commit a crime?
c. Will AI be liable for a crime? If not, who?

2. Defining Artificial Intelligence

In this disquisition, AI or an AI system is defined as a "a machine-based system designed to operate with varying levels of autonomy, that may exhibit adaptiveness after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments".[3] AI may be categorised into "those that (1) act like humans, (2) think like humans, (3) think rationally, and (4) act rationally."[4]

3. Criminal Law and AI: Can AI Commit a Crime?

Criminal law is an aspect of law which prohibits, criminalises and punishes acts.

[*] Francis Wayo is graduate of Law from Obafemi Awolowo University, Ile-Ife. His research interest in Artificial Intelligence (AI) interweaves the intersections of AI and intellectual property, human rights and environmental sustainability. He currently heads the Research and Publications of the Youth Centre of ILA-AI Committee on Artificial Intelligence, New Technologies and International Law, Nigeria. Contact: ilaaiyouthcentreresearchandpub@gmail.com .

[1] Gabriel Hallevy, When Robots Kill: Artificial Intelligence Under Criminal Law (Northeastern University 2013) xv.

[2] Ibid.

[3] European Union (Artificial Intelligence Act 2024, Article. 3 (1)

[4] Hallevy (n 1) 5

Acts which are deemed criminal or crimes are defined in statutes or laws as criminal law in most jurisdictions is underpinned by the principle of legality which encompasses the notion that a person shall not be punished for a crime unless such crime is defined in a law. The basic elements of determining whether a crime is committed are actus reus and mens rea. Thus, for something to be deemed a crime, there must be the actus reus (actual act) and mens rea (mental state or intention) on the party accused of the crime.[5]

While the perception of AI as a tool by human beings is a truism, the quest for machina sapiens as the holy grail of AI research has led to the development of thinking machines or computers that can think. These systems or machines can act independently without human directive, therefore, enabling them to engage in illegal acts such as "autonomous drone attacks, cyber-attacks based on machine learning, data poisoning, misuse of robots, and manipulation of stock markets".[6] Like the 1981 case involving the murder of an employee by an AI robot, there are other instances where AI controlled systems have committed acts criminal in nature. For instance, in April 2024, a Tesla car which was in "full self-driving mode" hit and killed a 28-year-old motorcyclist in Seattle.[7] Notably, the "full self-driving" feature is supposed to enable the car to run without any human control.

In theorising the practicability of AI crime (AIC), two experiments shed lights on the varying ways AI can commit crimes. The first is when an AI is used as an instrument to persuade "social media users to click on phishing links within mass-produced messages."[8] In this case, if a user clicked the link and later filled any form on the site, they would unknowingly be providing a criminal with personal data and information which can be used to commit a crimes. In the second experiment, it was found through a simulated

[5] Introduction to the Laws of Timor-Leste Criminal Law (Standford Law School <https://www.google.com/url?sa=t&source=web&rct=j&opi=89978449&url=https://law.stanford.edu/wp-content/uploads/2018/04/Timor-Leste-Criminal-Law.pdf&ved=2ahUKEwiisOXB9qaJAxW4U0EAHTnHHPg4ChAWegQIHBAB&usg=AOvVaw3vI0GKLOHHXyLqKiQ-hgVe > accessed 25 October 2024.

[8] Alaa Saud, 'Criminal Liability about the Use of Artificial Intelligence: Investigating the Actus Reus Element of AI-driven Technology' (2023) 6(1) American Journal of Law 1

[7] Hyunjoo Jin, 'Tesla car that killed Seattle motorcyclist was in 'Full Self-Driving' mode, police say' (Reuters, 31 July 2024) < https://www.reuters.com/business/autos-transportation/tesla-was-full-self-driving-mode-when-it-hit-killed-seattle-motorcylist-police-2024-07-31/ > accessed 24 October 2024.

[8] Thomas C. King, Nikita Aggarwall, Mariarosaria Taddeo and Luciano Floridi, 'Artifcial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions' (2019) Science and Engineering Ethics <https://doi.org/10.1007/s11948-018-00081-0 > accessed 25 October 2024.

process "that trading agents could learn and execute a "profitable" market manipulation campaign comprising a set of deceitful false-orders."[9] These two scenarios highlight two ways AI can be involved in a crime: when it is used by human as a tool and when it can act independently.

4. Can AI be held Liable for any Crime?

Gleaning from the above scenarios depicting possible situations where AI can be involved in a crime, who will be liable in such circumstances? In a situation where the AI was used merely as a tool by a person, it suffices to opine that liability will be on the person who orchestrated the criminal act or used the AI. However, what will be the case when the AI acted independently without human intervention? Will it suffice to hold the AI responsible?

In answering the above questions, a fundamental point worth mentioning is that the law prescribes who and who can be liable for a criminal offence. In Nigeria, for instance, the law recognises only human and juristic entities as legal persons capable of suing and being sued. This position has been given judicial flavour in the case of Management Enterprise Ltd v. Otusanya.[10] Since AI is not a person under the law, it is, therefore, not possible to hold it liable. Consequently, liability will have to shift to the developer, owner or/and user of the AI even when such persons neither nursed any criminal intention nor participated in any criminal activity.[11]

Furthermore, if AI can be held liable for a criminal offence, the mens rea and actus reus must be proved to establish the guilt of the AI. Establishing these elements may be practically cumbersome if not impossible. Since AI is not a person, proving culpable state of mind or guilty intent to carry out an act which turns out to be criminal may be very difficult. Unless an AI is developed and trained to facilitate or commit crimes, such systems are not inherently inclined to engage in criminal acts. While acts such as the killing of the factory worker by a robot and hitting of a cyclist by Tesla are examples of the actual occurrence or omission of an act which denote the actus reus of a crime, non-establishment of the mens rea may exonerate the AI from criminal liability as these elements need to be proved. Furthermore, finding AI liable will require determining its role in the offence as principal offender, an

[9] Ibid

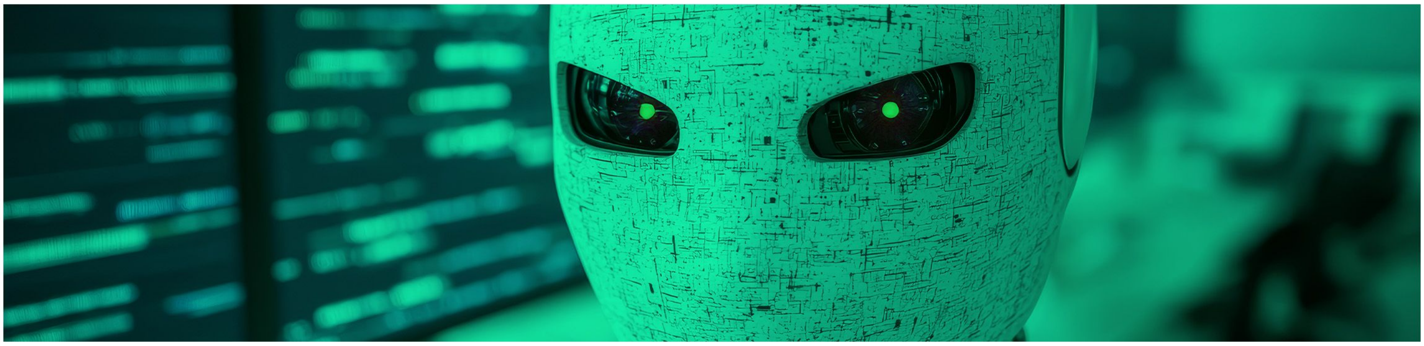
accomplice or accessory after the fact.

Due to the complexity in apportioning liability in crimes involving AI crimes, Hallevy proposes three legal models in dealing with the situation. The first method suggests holding the AI responsible by attributing both actus reus and mens rea to an AI system. He argues this position by referring to the fact that although criminal law was originally designed for humans, its scope has since expanded by applying to corporations, which are nonhuman entities.[12] But while it may be possible to assign actus reus to an AI, it will be difficult to prove the mens rea as mentioned earlier. Meanwhile, this theory may well apply to strict liability offences where intent to commit a crime is not necessary.[13] According to Hallevy, the second way of apportioning liability in AI crimes is via the perpetrator-through-another model. This is a type of criminal liability which is a form of vicarious liability. It has been defined as "aware execution of a criminal plan through instrumental use of another person, who participates in the commission of the offence as an innocent agent or semi-innocent agent."[14] Under this model, the AI developer/programmer, owner, or user is/are likely to be liable depending on whose initiation a crime is committed. In addition to the two, there is the natural-probable-consequence model which holds that a person is to intend the probable consequence of their action. This reasoning will also heap blame on the AI developer/programmer, owner, or user in the sense that they are expected to anticipate the consequences of their products. However, this will pose a difficulty where such act could not have been presumed or intended and the AI acts independently.

[12] Hallevy (n 1), p 67

[13] J.K.C. Kingston, 'Artificial Intelligence and Legal Liability' (2018) <https://www.google.com/url?sa=t&source=web&rct=j&opi=89978449&url=https://arxiv.org/pdf/1802.07782&ved=2ahUKEwj2kKSArq6JAxXPQkEAHWssMyoQFnoECBIQAQ&usg=AOvVaw0se-AgO0sPgmFNYKibMUWU > accessed 24 October 2024.

[14] Hallevy (n 1) p 69.



5. Conclusion

From the discussion in this paper, it is established that AI has the capability to commit crimes, and as the development of AI systems which can self-execute acts is on the rise, finding who to hold responsible is important. However, the extant criminal law system, which applies mainly to human beings and corporations, does not provide for a way which an AI can be held responsible as it lacks personhood and legal capacity to commit a crime and be sued. This then shifts liability to developers/programmers, owners and users of AI who may be susceptible to pay for the consequences which they did not intend nor anticipate as AI systems with capabilities of making autonomous decisions can act without any human intervention. Even if AI were to be held liable, the notion of culpability under present criminal law regime does not give much room for proving the culpability of AI in a crime. Against the found inadequacy, there is need for reformation of criminal jurisprudence to capture AI crimes. This is necessary to set parameters for determining liability in AI crimes and the penalties. While some have argued for granting AI legal personality like corporations, this author currently finds such argument unconvincing as personhood comes with rights, duties and liabilities which AI will not be able to meet in very many ways. Instead, the law can set parameters in apportioning liability to persons responsible for the AI.

Disclaimer: Views expressed in the article are solely of the author and do not represent the opinions of the Committee.