

Comparative Analysis of Criminal Liability of Autonomous Driving and Strong Artificial Intelligence

Artificial intelligence and its advanced technologies pose numerous threats and risks to both individuals and society at large, potentially leading to violations of human rights and criminal acts. The potential harm caused by AI systems can jeopardize the health and moral fabric of democratic communities. This research aims to explore the issue of criminal liability concerning strong and weak AI, focusing on who should be held responsible if AI systems cause harm or infringe upon the rights of individuals, groups, or society as a whole. Additionally, the study seeks to determine whether existing criminal law doctrines are sufficient to address AI-related offenses or if the need for a new specialized criminal law is imperative.

Product Liability

When it comes to AI, the well known principles apply. Manufacturers of autonomous cars must conduct rigorous testing based on the current state of science and technology to avoid liability for any damages or injuries caused by their products. It is their responsibility to continually monitor and improve their autonomous vehicles through measures like firmware updates or recalls, ensuring the safety and reliability of these technologies. Given the direct impact of autonomous cars on public safety and human lives, failure to fulfill these duties could result in criminal liability.

However, this approach may not be directly applicable to strong AI due to its nature as a self-learning system, potentially reaching or surpassing human intelligence levels. Unlike with autonomous cars, it might be impractical to hold producers indefinitely responsible for strong AI's actions. This challenge calls for a reevaluation of the traditional product monitoring and testing framework to accommodate the complexities of strong AI systems. Currently, no established standards exist for testing highly innovative technologies like strong AI, primarily because true Strong AI does not yet exist. Testing autonomous cars, though challenging, pales in comparison to the complexity of setting standards for testing strong AI.

If we were to draw inspiration from civil strict liability for criminal liability, it is important to recognize that there are cases where such an approach may not be suitable for holding individuals accountable for faulty products. While civil strict liability can offer valuable insights for certain aspects of criminal liability, it is crucial to acknowledge the fundamental differences between civil and criminal law, as their goals and considerations diverge significantly.

The general principle remains that anyone who introduces a risk should take all necessary measures to prevent harm to others, especially if the risk exceeds the basic risks inherent in everyday life. However, producers cannot be expected to anticipate and avoid risks that are unforeseeable given the current state of science and technology. Imposing such an obligation on producers to mitigate unforeseeable risks beyond the current knowledge could stifle innovation and impede technological progress.

Criminal law recognizes that uncertainty and unforeseeable risks are inherent in life and technological advancements. It accepts that absolute certainty and foresight are not always attainable and that risks are an integral part of human existence and progress.

In conclusion, diligent companies that fulfil their duty of care by thoroughly testing their products according to the state of science and technology will likely not be held liable for bringing faulty products to the market, provided they were not detectable after careful testing. Complying with obligations to conduct rigorous testing based on the current state of knowledge and technology can serve as a strong defence against liability claims for unforeseeable faults or defects that were not apparent at the time of product launch.

According to many jurisdictions, producers meet their duties by adhering to national and international licensing regulations, conducting risk assessments according to the current state of science and technology, and informing customers about potential risks. However, if a producer is aware that their product still contains errors likely to cause harm in the future, even after passing all relevant regulations, they may be held liable for damages. Releasing a product with known risks or defects without addressing them demonstrates negligence, as it knowingly puts consumers at risk despite being aware of the product's flaws.

Electronic Personhood

The introduction of electronic personhood presents a viable solution to address the issue of diffusion of responsibility. Legal personhood, as seen in the case of companies, involves the amalgamation of legal capacities, financial responsibilities, and material liabilities. While legal persons are treated as humans under the law in certain aspects, they do not enjoy the same legal status as human beings. However, it's worth noting that the category of legal persons does not encompass all groups, as it is a decision made by law to grant legal status to specific entities.

Corporate liability has proven to be a sufficient approach to hold individuals responsible for their actions within a company, while no one is held accountable for damages resulting from the company's activities. In fact, many countries have already established criminal liability for corporations.

A similar approach could be taken for autonomous machines. In theory, robots could develop an artificial personality and a certain scope of action and decision-making. This suggests that it is conceivable to create a legal status as a tangible representation for the collective cooperation of everyone involved in creating and using the robot. Internally, creating new legal entities with distinct legal obligations is not a significant challenge. However, it would require these autonomous machines to possess a certain degree of legal autonomy. Granting legal personhood to machines would be advantageous, as it would consolidate all legal responsibilities of the various parties involved.

This approach would have implications for civil law, as judgments could be directly issued against electronic persons. The judgments would be covered by the electronic person's assets, contributed by the parties involved in the creation and training process. In cases where a machine malfunctions due to a severe lack of care or intentional wrongdoing, the responsibility for payment could be attributed to one of the parties responsible for the machine's development or use.

However, it is essential to clarify that the concept of electronic personhood is not based on robots being considered artificial humans. To establish such a concept, it would need to be grounded in ontological factors that compare machines to humans, evaluating

traits like mobility, sensory perception, capacity for learning, intentionality, identity, capacity for reasoning, responsiveness to reasoning, ability to hold second-order desires, mental soundness, and other related attributes. The challenge lies in the fact that even human beings have not definitively fulfilled all the characteristics of consciousness.

The application of corporate criminal liability varies across different countries, with some, like Germany, opposing it. The academic debate is divided into two arguments: one suggesting that such a concept aligns with the inner logic of the legal system, while the other argues that there are insufficient similarities between corporations and humans, making criminal law inapplicable to them. This latter point gains significance when considering the even greater dissimilarity between corporations and robots. Robots directly interact with humans in the physical world and may display emotions and empathic reactions, potentially blurring the line between fictional entities and social actors.

However, these considerations don't necessarily mean that introducing criminal liability for robots is inherently wrong. At present, given the state-of-the-art in AI, the non-similarity approach seems more convincing. This means that to introduce new entities like robots into domestic criminal law, significant changes would be required.

Are Changes in Criminal Law Needed?

Through an extensive examination of relevant literature, legislation, and judicial rulings on the criminal liability of AI, it is evident that the question of whether amendments to criminal law are necessary due to advances in AI remains largely unresolved. Nevertheless, this research sheds light on the potential implications for the development of criminal law, particularly regarding the direction of AI's criminal liability.

As demonstrated by this research, there exists a notable distinction between the criminal liability of weak and strong AI. Hence, it becomes essential to consistently differentiate between the two. Additionally, addressing the challenges associated with identifying the responsible party in cases of AI malfunction is of utmost importance.

The research in this thesis establishes that applying current criminal law to weak AI is not problematic. However, the primary challenge in the field of AI's criminal liability lies in identifying the responsible individual in case of a malfunction. Manufacturers are

generally considered to bear the primary responsibility for ensuring their AI systems' proper functioning. However, the complexity of assigning liability in criminal cases involving AI demands a detailed investigation of all relevant parties, including producers, programmers, users, and other parties directly involved.

“Blindly” assigning liability in situations like an autonomous vehicle failing to brake and causing a pedestrian's death is not straightforward. Thorough and meticulous investigations are necessary to determine the accountable party for an AI malfunction. Developing appropriate methodologies to accurately identify the responsible party becomes crucial in addressing this challenge.

Once the responsible individual is identified, implementing criminal law presents no significant challenges, as mentioned earlier. The focus should be on determining culpability for the malfunctioning AI system, rather than contemplating legal alterations to impose criminal liability on an individual.

The research demonstrates various liability scenarios for weak AI that are already well-established and applicable in areas like autonomous driving and other applications of weak AI. The need for legal adjustments should concentrate on creating guidelines for competent courts to effectively handle AI-related accidents and determine appropriate criminal liability.

Given the diversity of potential criminal deeds involving AI, no single existing legal model for liability can adequately allocate and impose criminal liability in all conceivable situations or the negative consequences arising from AI usage. Therefore, lawmakers and legal experts must continually analyse and adapt legal frameworks to keep pace with technological advancements and ensure proper liability allocation.

The issue of strong AI, which possesses intelligence at least equal to that of a human being, presents unique challenges in the context of criminal liability. Strong AI, like humans, is not entirely predictable in its decision-making behaviour, making it difficult to determine the predictability of its potential misconduct for the producer or any other party involved. Imposing criminal liability on the producer for the actions of strong AI may discourage the

development of highly intelligent machines capable of learning and adapting due to their unpredictability.

A thought experiment comparing strong AI to children highlights the potential injustice of holding producers accountable for strong AI's actions. Just as parents have certain obligations and liability for their children's actions until they reach a certain age, making producers liable for the unpredictable behaviour of strong AI could have significant implications for technological advancement and innovation.

Currently, strong AI lacks the prerequisite of personhood required for criminal liability. As a result, if strong AI were to exist today, it would not be subject to criminal liability for its actions. However, the fact that AI exists and has been attributed personhood raises important philosophical and legal debates surrounding the definition of personhood and its implications for AI ethics and responsibility.

One potential approach is to consider a framework similar to corporate criminal law for addressing strong AI within the realm of criminal liability. However, this would require substantial amendments or new legislation, as existing corporate criminal law may not cover transgressions like manslaughter or homicide involving AI. The need for innovative legal solutions that effectively address the unique challenges posed by AI in criminal liability is evident, calling for the development of specialized criminal laws (i.e., *lex specialis*) designed specifically for AI.

In conclusion, the emergence of strong AI raises complex questions about accountability, unpredictability, and the implications of attributing personhood to AI. Striking a delicate balance between accountability and fostering innovation in the field of AI is crucial. Policymakers and legal experts must navigate these challenging issues to develop appropriate legal frameworks for AI's criminal liability.

The second option is to treat AI under criminal law similarly to human beings. This approach is supported by compelling arguments. When holding humans criminally liable, there is no concrete scientific evidence for the existence of consciousness, and yet, criminal law manages to bypass the need for actual proof of any mental element. The mere awareness

of the outcome of their actions, or foreseeability, is sufficient for criminal law to presume the mental element or intention in humans.

Similarly, intention can be imputed based on knowledge of wrongful acts and a deliberate decision to proceed with the action for artificial intelligence. AI acquires factual data through sensors, just as humans use sensory organs like eyes and ears to gather information. In machines equipped with strong AI, all the data is processed in a central unit, leading to the generation of a final image from the factual data, much like how the human brain consolidates information.

These similarities in the acquisition and processing of factual data suggest that AI can exhibit mens rea, or the mental element required for criminal liability. This supports the argument for subjecting AI to existing criminal law and applying it in the same way as it is applied to humans. The differences between humans and strong AI may not be as significant as expected, making it unnecessary to distinguish between the two in terms of criminal law and culpability.

However, despite these arguments, it is acknowledged that the current state of technology and the rapid pace of advancement means that the criminal culpability of strong AI is likely to remain a thought experiment in the foreseeable future.

In conclusion, the idea of subjecting AI to criminal law similar to human beings raises interesting points about mens rea and the similarities in data acquisition and processing. While it is a thought-provoking concept, the practical implementation of criminal liability for strong AI remains uncertain given the current technological limitations. As AI technology progresses, ethical and legal considerations surrounding AI's criminal liability will continue to evolve.

Final Conclusions

This thesis delivers a comparison of the criminal liability between strong and weak AI. The results indicate that there are significant differences in the criminal responsibility of these two types of AI. Weak AI lacks cognition, intelligence, and awareness, making it merely a tool that humans use for various tasks, and thus, it cannot be subject to criminal law. In contrast, strong AI exhibits cognition, intelligence, and a level of awareness that

allows it to make conscious decisions, meeting the fundamental requirements of criminal law. Therefore, under current criminal law doctrine, strong AI could be held responsible in criminal courts, whereas weak AI would require a human individual to be held liable for its actions.

The thesis explored various legal bases on which a person can be held criminally responsible for crimes caused by weak AI malfunctions. Due to the many possible individuals behind weak AI, such as producers, programmers, and suppliers, the criminal court would need to navigate a complex network of possibilities to determine guilt. In contrast, strong AI is likely to operate without a specific individual behind it, making the process of determining responsibility less convoluted.

The study reveals that even with thorough testing and observation of weak AI products, criminal liability could still arise, placing companies, suppliers, and programmers at risk. The thesis suggests that guidelines for competent courts on handling such cases would be more appropriate than legislative adjustments in criminal law.

To support its findings, the dissertation draws from a diverse range of academic sources, including legislations, court decisions, commentaries, journals, peer-reviewed articles, scientific conference presentations, textbooks, online articles, and interviews.

In conclusion, the thesis clarifies that the criminal liability of weak and strong AI differs significantly, with the former requiring human accountability and the latter being potentially responsible in its own right. The study underscores the need for guidelines in handling cases involving weak AI, considering the numerous individuals involved, to ensure fair and appropriate legal outcomes.

The title of this thesis suggests a comparison of the criminal liability of strong and weak AI. However, the results demonstrate that the criminal responsibility of weak and strong AI is fundamentally different. Weak AI lacks cognition, intelligence, and awareness, making it incapable of being a subject of criminal law. As weak AI is used as a tool to support or perform tasks autonomously, the human behind the weak AI will always bear criminal responsibility.

The study shows various legal bases on which a person can be held criminally liable for a crime caused by a malfunction of weak AI. The similarities in liability lie not between weak and strong AI but between the human producer (or the respective party responsible for weak AI) and strong AI itself. Strong AI possesses cognition, intelligence, and awareness to a degree that allows it to make conscious decisions, meeting the fundamental requirements of criminal law. Thus, under current criminal law doctrine, proving mens rea for strong AI would not be an obstacle, making it responsible before criminal courts, unlike weak AI.

The thesis demonstrates that due to the various possible individuals behind weak AI, each of whom can be held criminally liable on multiple grounds, criminal courts must navigate a complex network of possibilities to find a person guilty. This complexity is not present in the case of strong AI, as no one is typically behind it. The results indicate that producing weak AI systems, even with thorough testing and observation, may not be entirely risk-free for companies, as criminal liability could arise. Adjustments in criminal law may be necessary, but not legislative ones; rather, guidelines on how competent courts should handle such cases.

21.08.2023

Nmas