The Liability of Artificial Intelligence's Moral Dilemma

[version 1; peer review: awaiting peer review]

Julia Farhana Rosemadi, Saidatul Nasuha Jamaludin

Faculty of Law, Multimedia University, 75450 Bukit Beruang, Melaka, Malaysia

Abstract

Background: Artificial Intelligence (AI) represents the fundamentals of the Fourth Industrial Revolution. Its advancement in technology has brought benefit to mankind over the years, which helps to amplify our daily lives in numerous ways. Among the technologies that have been innovated utilizing AI are Artificial Intelligence Lawyers (AI Lawyers), autonomous vehicles, AI judges, augmented drafting services and delivery drones. Although AI has brought many advancements in our lives, concerns are raised with regards to the moral implication of AI. This study aims to examine the situation wherein tools which incorporate AI would have to face a moral dilemma situation, or situations where AI had acted beyond what was instructed or programmed by the manufacturers. Thus, this work discusses whether the torts theory of liability on negligence is applicable in such a context. Methods: The methodology involved a doctrinal research in analysing the legal rules and existing studies in relation to AI and the negligence principle to develop a critical analysis of the literature. Besides, this research finds the application of principle negligence adopted in other countries, particularly in the United States, since it has employed AI much earlier compared to other countries. Results: The findings suggest that the question of the liability of AI is still at its infancy, as policy makers or regulators have yet to decide the liability of AI. However, the findings also suggest that theories of liabilities such as product liability and strict liability are possible to adduce liability on AI, so long as AI could be defined as a product. The research limitation is the scarcity amount of related studies available in the literature. The motivation of this work is in the importance to determine whether advanced AI may be liable for ‘moral decisions’ that cause damage to the victim under the principle of negligence.

Keywords
Artificial Intelligence, Ethical Issues, Moral dilemmas, Negligence, Liability
This article is included in the Research Synergy Foundation gateway.

Corresponding author: Julia Farhana Rosemadi (julia.farhana@mmu.edu.my)

Author roles: Rosemadi JF: Conceptualization, Methodology, Supervision, Writing – Original Draft Preparation; Jamaludin SN: Formal Analysis, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: The author(s) declared that no grants were involved in supporting this work.

Copyright: © 2022 Rosemadi JF and Jamaludin SN. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Rosemadi JF and Jamaludin SN. The Liability of Artificial Intelligence’s Moral Dilemma [version 1; peer review: awaiting peer review] F1000Research 2022, 11:1079 https://doi.org/10.12688/f1000research.73640.1

First published: 21 Sep 2022, 11:1079 https://doi.org/10.12688/f1000research.73640.1
Introduction
What is Artificial Intelligence?
In the wake of the fourth industrial revolution, Artificial Intelligence (AI) has paved its way towards helping mankind.
AI is defined as the ability of computer systems to perform tasks linked to human intelligence such as speech recognition.\(^1\) AI is also capable of making decisions and is used to assist decision-making, as it is inevitably taking over the current market.\(^2\) According to McCarthy, AI is defined as “the science and engineering of making intelligent machines, especially intelligent computer programs.”\(^3\) Since these technologies are more reliable and intelligent compared to humans, it is important to determine whether AI may be liable for ‘moral decisions’ that cause damage or injury to the victim under the principle of negligence.

Methods
Doctrinal research was adopted in analysing the legal rules and existing literature concerning AI, specifically that which focuses on the negligence principle. This research aims to develop a critical analysis on whether negligence may be used as the basis of liability for AI when making ‘moral decisions’. Besides, this research involves the application of principle negligence adopted in other countries, particularly in the United States, since it has employed AI much earlier than other countries.

Discussions
The moral dilemma in Artificial Intelligence (AI)
AI technologies are forced to make decisions and judgments. AI employs machine learning whereby it empowers machines to learn through experience and solving problems by employing predefined algorithms.\(^4\) These decisions and judgments are based on a set of algorithms that are incorporated into the system or programme. There is little human supervision, and AI is free to make its own decisions.\(^4\) By making decisions, AI is exposed to data. It requires an adequate amount of correct and non-biased data, which is a challenge in the real world.\(^5\) There is risk that an AI system would modify the real-world data according to its machine learning understanding, which might be biased or contain false information.\(^5\) Hence, increasing concerns are raised about the moral decisions made by AI. The following are several examples of the moral dilemma faced by AI technologies:

a) Autonomous vehicles

Autonomous vehicles are self-driving cars that can guide themselves without human intervention. Even though decision-making in autonomous vehicles involves technical algorithms, the judgment itself may entail some moral philosophy depending on the manner in which it is programmed.\(^6\)

b) ‘AI lawyers’ and augmented drafting services

In 2016, the first AI lawyer known as ROSS was employed in one of the largest law firms in the United States.\(^7\) ROSS utilizes machine learning capabilities and mainly collects passages from cases and legal sources in order to answer legal questions.\(^7\) Thus, asking questions to ROSS is similar to asking questions to a real lawyer. A question arises on whether ROSS is capable of performing moral judgments that are in line with the law. Possible liability for AI, aside from negligence, is where ROSS makes a decision that goes beyond moral values or against the ethical code. Imagine a situation where ROSS has falsely reviewed a contract that may jeopardize a client’s position. In certain cases, the moral decisions made by AI lawyers or augmented drafting services may thus have a negative impact on society.

c) AI judge

An AI judge would analyse the sources of law and make decisions based on its findings.\(^8\) This gives judgment based on algorithms found on legal databases.\(^8\) In State v Loomis, a man was being sentenced to six years imprisonment by using an algorithm software called COMPAS.\(^9\) COMPAS is a software where the system analyses a person’s data and makes recommendations to the judges on the length of the sentence in prison a person should get.\(^9\) In order to explore the pattern of judgments, decisions from courts are analysed by using a case-to-case basis.\(^9\) However, an AI judge was being criticized in the State v Loomis, as the system was racially biased, since it predicted that black men are future criminals two times higher compared to white men.\(^9\) Therefore, a question arises on whether the decision of AI judge is always moral and reliable.
The question of morality in AI

In designing AI, the trolley problem is always posed in situations where AI is required to make important decisions. This can be illustrated by imaging a situation where a trolley is charging towards five workmen on the track. The trolley is not able to stop due to defective brakes. A person is standing next to a switch where he can divert the trolley to another path. However, a single workman is present on another path. Thus, if that person pulls the switch to another path, the single workmen will die. On the other hand, if that person does not pull the switch and revert the path, five workmen will die. Similar to AI, the person who holds the switch is the programmer. The difference is that the programmer is not a bystander. Therefore, his actions will bring about consequences, since he has to make decisions. However, the trolley problem does not solve the question of liability, as the trolley problem only considers how a vehicle should react or make moral decisions.

Liability arising out of the moral dilemma of AI by applying principle of negligence

Liability is often questionable when it comes to AI. Based on the scenarios discussed above, if the victims suffer any damage, they may seek recourse under the existing liability regime through the tort system, to claim compensation under the principle of negligence. The negligence principle is recognised in the United States as well as in common law countries. In tort, the injured party will seek redress from the negligent conduct of the defendant. The damages will be paid by the person causing the damages. Thus, AI companies are always on the lookout, because these companies are the ones who have the best financial capabilities to pay damages.

This research seeks to discuss whether the principle of negligence may be used to claim damages that arise out of the ‘moral decisions’ of AI. It is pertinent to discuss the principle of negligence first, before analysing its basis for liability. Negligence in tort is defined in the case of Blyth v Birmingham Waterworks Co as “the omission to do something a reasonable man would do, or doing something which a reasonable man would abstain.” To be held liable under negligence, several elements need to be satisfied: the duty of care, breach of duty, causation, and damage. Hence, to assess whether AI may be held liable under negligence, these elements are discussed in detail.

Duty of care

Under the element of duty of care, the classic case used under common law is the case of Donoghue v Stevenson, wherein reasonable care must be taken to avoid acts or omission that are likely to injure the person closely and directly affected by the defendant’s act. The duty of care was later extended in the case of Caparo Industries v Dickman and Others, where the court held that the harm must be foreseeable, there must be proximity between the plaintiff and defendant, and it must be fair and just to impose a duty of care. In determining the duty of care concerning AI, it might be less compared to the duty of care that does not involve AI.

Other than that, since there are multiple parties involved in producing AI, a suitable question to ask is: who owes a duty of care towards whom? The number of parties involved may range from programmers, to engineers and manufacturers. Another question that arises is whether the traditional reasonable man’s test as applied in common law to determine negligence applies to AI. Similarly, in the United States, the standard of care is assessed using the reasonably prudent person. Since there is an absence of a particular standard for liability under negligence, an alternative of the standard of care, aside from the reasonable man’s test, may be applicable. A professional standard of care may be appropriate to be imposed, as the persons who created the software are regarded as professionals.

Breach of duty

Breach of duty may arise in several situations, as follows:

i) A breach may occur due to malfunctions or errors in the software.

ii) A breach may also occur if the vendors have entered wrong information given by experts, or poor design or maintenance by the manufacturers.

Hence, establishing a breach is not easy, as there are many parties involved, especially when AI involves advanced technology. To prove liability involving multiple parties is not easy either, especially when AI adopts machine learning.
Causation

Causation and foreseeability is the main ingredient to establish liability under the tort of negligence. These issues must be confronted before establishing damage due to decisions made by AI, under the tort of negligence. To succeed in a negligence claim, it is material for the plaintiff to prove that there is a reasonable connection between the act of the defendant, and the damage caused. Thus, the challenge in proving the cause of the damage, or causation, is the multiple parties and the complexity involved in the programme, making it difficult to pinpoint the originating cause of damage.

Furthermore, in establishing causation, the damage must be foreseeable. If the damage is unforeseeable, negligence cannot be established. Thus, if AI reacts by making ‘moral decisions’, the damage in consequences of the act must be foreseeable. If there is a lack of foreseeability in AI making decisions, there is a possibility that liability under negligence would not be blamed on anyone at all. According to Yavar Bathaee, the law “is built on legal doctrines that are focused on human conduct, which when applied to AI, may not function”. This is because it is impossible to determine how AI would come up with the decision through internalized learning of a black box nature, hence making its decision unforeseeable. It can be illustrated by the ‘moral decisions’ discussed, where an autonomous vehicle would fully recharge itself the whole night because the vehicle discovered by machine learning that it would perform better with a full battery. However, for the autonomous vehicle to fully charge itself, it releases carbon monoxide and thus endangered the household. The manufacturers may claim that it is foreseeable that autonomous vehicles would get into an accident. However, they never foresee that autonomous vehicles may kill anyone by the release of carbon monoxide. Thus, to develop a balanced law, the producers of AI must take precautions to evaluate the foreseeable risks of AI technology.

In assessing the elements of negligence, it is difficult to establish liability under the tort of negligence, due to the issues discussed above. Thus, it is a challenge to create AI that employs machine learning to be liable under the tort of negligence, since it seeks the negligence of humans, not machines.

Possible liability for AI aside from principle of negligence

Product liability

Aside from negligence, product liability may also be considered to impose liability under the existing liability regime, as it may occur when there is a defect in the product manufactured. In the United States, product liability claims may also be brought under negligence, strict liability or breach of warranty. Since humans are not involved when AI robots make decisions based on machine learning, liability may move to product liability. Under this principle, it must be construed whether the algorithm used in machine learning is considered an actual product. There is no clear discussion on the status of the algorithm as a product. In determining the liability of AI by using the product liability principle under Malaysian law, the first crucial issue to be determined is whether the algorithm used in AI machine learning is recognized as a product or a service. If such an algorithm is determined as a product in AI, then the product liability principle can be applied to prove the liability of AI.

Strict liability

Another option to assign liability to AI is to adopt strict liability. Under strict liability, the manufacturer would be held liable, regardless of whether or not he is at fault. There is a possibility that manufacturers may be held liable for the decisions made by AI robots. However, it must be noted that if there are changes to the product after delivery, manufacturers may not be held liable under strict liability, unless the changes made to the product are foreseeable. Hence, concerning AI products that employ machine learning, the element of foreseeability must be construed to make manufacturers strictly liable. The absence of foreseeability when AI causes harm when using machine learning may lead to manufacturers not being liable.

AI personhood and claims pool

Instead of assigning liability to manufacturers or other parties, the conferment of AI personhood would make AI independently liable, where it is capable of being sued in court. Another option is to make it obligatory for AI to be insured or have a universal insurance claims pool to pay out damages. A no-fault liability scheme may be established by contributing to a claims pool to pay for damages resulting in harm by AI robots, where the claims pool may be contributed by the AI and robotics industry.

Conclusions

Based on the discussion above, the test of whether the torts theory of liability on negligence is applicable in such situations is answerable, as a claimant intends to use negligence solely to produce a civil claim against AI is insufficient. This should
be supported by strict liability principles and product liability principles to prove such liability. Alternatively, treating AI as personhood would give room for AI independently being liable and capable of being sued in court.

The complexity of this issue is that the current law may not be able to withstand the challenging legal issues on AI. Since there is an absence of a clear legal framework in deciding on AI liability, it is a great concern and responsibility for stakeholders to draft regulations regarding the liability of AI. It is the right timing for stakeholders to propose a regulation or set of laws not limited to aligning the objectives of AI and its functions, roles, elements, characteristics and liability. Nevertheless, theories of liabilities such as negligence, product liability and strict liability are possible to adduce liability on AI. It should also be noted that the liability of laws is successful to be proved against AI, only if it is considered a product.

In conclusion, AI brings many benefits to mankind. However, it does come with a new set of problems. Since AI robots may employ machine learning, they are capable of making their own decisions, which may lead to harm. In discussing the theory of negligence as a basis for liability, it was found that persons who produce machine learning AI that has resulted in harm may escape liability, mainly due to lack of foreseeability. Hence, determining liability concerning AI which employs machine learning remains a challenge, and merits further exploration, especially since AI is inevitably becoming more intelligent with time.

**Author contributions**

The research was designed by JFR. JFR carried out the research by compiling the materials and drafted the manuscript on the part of introduction, methodology, discussion and conclusions. SNJ carried out the research by compiling the relevant materials and help to draft the manuscript concerning product liability as well as reviewing the manuscript. Both authors read and approved the manuscript.

**References**

The benefits of publishing with F1000Research:

• Your article is published within days, with no editorial bias
• You can publish traditional articles, null/negative results, case reports, data notes and more
• The peer review process is transparent and collaborative
• Your article is indexed in PubMed after passing peer review
• Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com