

Rewriting Delictual Liability in the Age of AI: Assessing Negligence for Physical Harm Caused by AI Driven Robots in Sri Lanka

Didula Dharmawardhane^{1*}, Dr. Samurdhi Jayamaha²

¹School of Law, Sri Lanka Institute of Information Technology, Colombo 03, 00300, Sri Lanka

²School of Law, Sri Lanka Institute of Information Technology, Colombo 03, 00300, Sri Lanka

Corresponding author *: diduladharmawardane@gmail.com

Abstract

Where an Artificial Intelligence driven robot causes physical injuries to human beings, Aquilian action under the law of delict shall be applicable by default considering that there are no statutes or case law governing such incidents. Nonetheless, the application of the traditional delictual law doctrines in this context is difficult or rather impractical due to their inherent characteristics such as lack of explainability, unpredictability, autonomy, and multi-party involvement. The objective of this paper is to analyse the aforesaid issue and provide recommendations to resolve the matter. While the paper indicates how conventional Aquilian action fails in this context, it offers recommendations to incorporate recent developments in the field of AI and judicial/ legislative requirements.

Keywords: Artificial Intelligence; Advanced Robots; Delictual Liability; Aquilian Action

1. Introduction

The rapid and continuous advancement of Artificial Intelligence has given rise to a range of legal challenges that existing legal doctrines are not adequately equipped to resolve. One such instance arises when advanced robots powered by artificial intelligence, particularly those designed to interact with and, in some cases, collaborate alongside humans, cause physical injuries to individuals. The objective of this research paper is to analyse the suitability of the Aquilian action, under the current delictual liability framework of Sri Lanka, to address cases involving Artificial Intelligence-powered advanced robots (hereinafter 'AI-robots') and to provide recommendations for any potential legal reforms in this regard. The paper primarily adopts a doctrinal, library-based methodology. Part I delineates the subject matter and discusses four inherent attributes of such AI-robot which materially impact the application of delictual liability in a technical perspective. Part II proceeds to analyse the implications of these attributes for the application of the Aquilian action in delict (*actio legis Aquiliae*) and subsequently offers recommendations.

2. PART I: Attributes of Artificial Intelligence Powered Advanced Robots

Consider an example where an AI robot accidentally injures an individual due to a mistake. The incident may have happened due to multiple reasons. In such circumstances, it becomes difficult for a human observer to ascertain the precise cause of the incident, as it is not readily apparent which input data the robot utilized in its decision-making process, nor how those inputs were evaluated and weighted across the numerous hidden layers within its neural architecture¹. In Artificial Neural Network, decisions are made through millions of weighted connections across multiple hidden layers. The outcome is determined by the connection with the highest predicted probability. Human observers usually perceive only the input variables and the output but the internal process or intermediate connections (process within the hidden layers) that led to the decision remain largely unclear². This lack of transparency resulting from ‘the complexity of large multi-layered neurons’ is known as the ‘black box’ phenomenon. The black-box phenomenon challenges the interpretability, explainability and traceability of AI- Robot.

In the primary stage of AI where supervised training takes place, the developer is aware of the outcome. However, Learning utilized in advanced robotics is mostly ‘unsupervised’ and ‘reinforcement learning’ algorithms, hence the robot is left at its discretion to decision making through input data. This phenomenon results in a lack of predictability³. In fact, the more advanced the AI is, the more complex it is to predict the outcome for any party. involved. Predictability is an intuitively familiar concept where one can anticipate the outcome of a certain process without understanding its underlying mechanism. The programmer cannot anticipate, in precession, the output of an advanced robot out of millions of probabilities calculated within its processor. In fact, advanced AI is capable of exhibiting behaviors that are unexpected, and the rationale for this is that AI develops novel behaviors that were not programmed as a result of its interaction with the world and other agents⁴. Therefore, the parties involved could not have reasonably anticipated caused by the AI-robot to the human in advance.

In the context of AI, autonomy refers to its cognitive process which allows decision making without external inference and through its own learning from the input data similar to that of a human’s thought process *albeit* notionally. Integration of ML and particularly DL to AI-robotics have illustrated their increased cognition to learn from experience. Barfield, for instance, quotes an example in his work on how ‘Baxter’ an industrial robot, learns to grasp objects through deep

¹ Woodrow BARFIELD, “Liability for Autonomous and Artificially Intelligent Robots” in Paladyn Journal of Behavioral Robotics, 2018, Volume 9, Issue 1, pp. 193-203

² Rachum-Twaig, Omri, Whose Robot Is It Anyway?: Liability for Artificial-Intelligence-Based Robots (February 21, 2019). University of Illinois Law Review, Vol. 2020, Forthcoming, Available at: SSRN: <https://ssrn.com/abstract=3339230> p.7-8.

³ Benhamou, Yaniv & Ferland, Justine. (2020). ARTIFICIAL INTELLIGENCE & DAMAGES: ASSESSING LIABILITY AND CALCULATING THE DAMAGES.

⁴ Yampolskiy, Roman. (2020). Unpredictability of AI: On the Impossibility of Accurately Predicting All Actions of a Smarter Agent. Journal of Artificial Intelligence and Consciousness. 07. 109-118. 10.1142/S2705078520500034.

learning⁵. Learning can be differentiated into two parts: self-learning and continuous learning. Former refers to the AI-robots ability to recognize patterns in training data while latter refers to the incremental learning of AI during continuous operations. Both self-learning and continuous learning allow the robot to take decisions and adapt to an environment. Therefore, it can be argued that the more autonomy machines achieve, the more tenuous the strategy of attributing and distributing legal responsibility for their behavior to human beings becomes⁶. Accordingly, autonomy interferes with the ability of human actors to exercise full control over AI-robot's actions.

Consider the AI robot, which injured the human is a collaborative industrial robot. Nicole Brex et al observe that in a collaborative industrial robot (similar to the robot in example) four stakeholder groups contribute ; 'i) system actors (the product/service users); ii) system experts (professionals such as engineers and system designers); iii) system decision-makers (e.g. managers who decide about the requirements for the system design, the purchasing, the implementation, and the use of the system); and iv) system influencers (e.g. media, standardization organizations, regulators, and citizens)⁷. Therefore, AI-robot itself is a product of multi-party involvement. Nonetheless, Nicole's classification focuses on broader picture and is not suitable for the objective of this paper. Henceforth, the paper adopts a focused approach encapsulating the ones who create, own, and deploy which *inter alia* include i) Developers (who design and manufacture the robot), ii). Operators (who operate the robot) and iii.) End-Users (who utilize the product)⁸.

3. PART II: Delictual Liability and Physical Injuries Caused by AI-Driven Robots

In continuation of the earlier illustration, consider a situation wherein an autonomous factory robot accidentally struck a worker while moving materials, resulting in grievous bodily injury due to the robot mistakenly identifying the worker as an inanimate object. In a liability-driven discussion on AI, one must consider the possibility of imposing liability on the AI itself, whether by recognizing it as a legal person (natural or juristic) or by applying vicarious liability, assuming it functions as an agent, due to its autonomous nature. Nonetheless, it appears premature to impose liability directly on AI robots, as the parties involved still retain a significant degree of control over their design, deployment, and functionality, although, in the foreseeable future, it may become necessary to re-evaluate this position. Nevertheless, this paper does not seek to engage in an in-depth discussion of this issue. Rather, it proceeds with a presumption that the parties who created, owned and deployed the AI entity shall bear the primary liability for their

⁵ Barfield, Woodrow. "Liability for Autonomous and Artificially Intelligent Robots" Paladyn, Journal of Behavioral Robotics, vol. 9, no. 1, 2018, pp. 193-203. <https://doi.org/10.1515/pjbr-2018-0018>

⁶ Chinen MA. The co-evolution of autonomous machines and legal responsibility. Va J Law Technol. 2016;20(2):338-393.

⁷ Berx N, Brescia A, Aqamarina R, Curcio EM, Pintelon L, Carbone G. Stakeholders' perspectives on safety related human-robot collaborative scenarios. International Journal of Advanced Robotic Systems. 2023;20(5). doi:10.1177/17298806231200095.

⁸ Rachum-Twaig (n.2).

negligence provided that wrongful conduct or omission on the part of defendant(s) is not observed.

The foremost requirement of Aquilian actions, Wrongfulness is objectively assessed *ex post facto* by considering whether a legally recognized right was infringed causing a harmful result⁹ and whether it occurred in a legally reprehensible manner. An act is not regarded as delictually wrongful unless it results in consequences or harm. The former is *prima facie* established considering the robot does infringe the legally recognized right to bodily integrity (personal right) of a human being which resulted in the consequences of physical injuries. The latter, however, is rather difficult to be attributed to the defendants. The *boni mores* criterion consider whether the legal convictions of the community regard the act or omission as wrongful for the purpose of delictual liability¹⁰. If the developer failed to train the AI model to interact properly with human beings, or if the operator failed to follow instructions (for instance, by operating it in a busy area), or if the end-user interfered with the robot's functions (in the case of an incident involving a third party), it could be argued that this resulted in a human being injured, and thus constitutes a delictually wrongful act.

However, the AI robot's autonomy allows it to take actions without the direct knowledge or oversight of the human parties involved. In such cases, it may not be possible for the developer, operator, or end-user to predict the AI robot's decisions and thereby take the necessary precautions to minimize or eliminate potential consequences. In that context, can the community consider the act of manufacturing, deploying, or utilizing the AI robot as the cause of the consequence, and thus regard it as a delictually wrongful act provided that there has been no error by way of act or omission on human's part? Therefore, the paper argues that wrongful act should be inferred on the basis of the nature of harm and control over the risk. Lemley and Casey analyse the range of harms that is to be expected from an autonomous robot such as unavoidable harm, unforeseen harm, misuse harm and defect -driven harm¹¹. While some of these harms are avoidable by the operators and/or developers, there are harms associated with its unpredictable nature or unexpected circumstances.

For example, a robot carrying a heavy load in a children's playground may collide with a child who suddenly enters its path, due to its inability to alter trajectory in time. This constitutes unavoidable harm, as the outcome would not differ if a human were in the robot's place. Alternatively, if the robot's actions resulted from its own learning through experience, the harm would be classified as unforeseen harm, since neither the developer nor the operator could have reasonably anticipated the conduct. In both cases, while harm occurred, no delictually wrongful act was committed by the human parties due to the absence of foreseeability and control. Conversely, if the robot was deployed in a playground despite clear warnings against human interaction, the operator commits misuse harm. Likewise, if the robot was inadequately trained

⁹ Pelis And Another V. Arnaashal (1980) 2 Sri L.R. P.298

¹⁰ Coronation Brick (Pty) Ltd v Strachan Construction Co (Pty) Ltd 1982 4 SA 371 (D) 380

¹¹ Lemley, Mark A. and Casey, Bryan, Remedies for Robots (July 31, 2018). 86 University of Chicago Law Review 1311 (2019), Stanford Law and Economics Olin Working Paper No. 523, Available at SSRN: <https://ssrn.com/abstract=3223621> or <http://dx.doi.org/10.2139/ssrn.3223621>.

despite prior notice of its intended use near children, the developer incurs defect-driven harm. In these latter cases, the harm is foreseeable, control is sufficient, and a wrongful act is clearly attributable to the responsible human party.

Once the wrongful act or omission is justified, the patrimonial loss should be established by the plaintiff, something which shall be capable of assessing in terms of money. Under an exception to this rule, compensation may be awarded for pain and suffering in respect of personal injuries¹². In the context of physical harm caused by an AI robot, the resulting patrimonial loss, such as medical expenses, lost income, or diminished earning capacity, can typically be established with relative ease, given the tangible and measurable nature of the damage.

Subsequently, the fault shall be established in terms of *Culpa* or rather negligence, a conduct of unintentional carelessness which resulted in harm. To assess this standard, courts apply an objective criterion known as the *diligens paterfamilias* (the reasonable person test), taking into account various factors¹³. When applied, the court's evaluation of whether the risk of harm was foreseeable to the extent that the claimant was likely to be injured and whether precautions could have avoided the damage, unless conduct is justified by social utility or customary practices, in fact cannot be entertained in the context of AI-powered advanced robotics. Practically speaking, as stated previously, neither developers nor operators can foresee the outcomes or the extent of the risks associated with the robot once deployed, owing to its intrinsic unpredictability, which further undermines their ability to implement critical precautions. Hence an inference that 'a reasonable, prudent individual can definitively anticipate the outcome' may sound absurd.

While it may be impossible for them to foresee robot's conduct technically, factually it's possible to do so. For instance, if the robot was deployed in a children's playground without adequate precautions, there's a reasonable possibility of an incident on the part of the operator. If the model is trained using biased, inaccurate data, then there's a possibility of an incident on the part of the developer. Nonetheless, asserting the standard on a case-by-case basis depending on the objective components may lead to inconsistencies thereby diminishing the appropriate standard which should be expected from system experts. Analogically, if the court were to infer a test similar to that of *Bolam*, *Bolitho* (pertain to medical negligence), the absence of a unified set of professional standards for specifically operator and developer may hinder the progress. Therefore, it's recommended that implementing a set of practices for developers & operators may justify the reasonable man standard expected.

Subsequently, the victim (claimant) shall demonstrate that such breach caused them an actionable loss. Accordingly, the courts ask: Did the defendant's negligent act actually bring about the harm suffered by the plaintiff? (factual causation), and to what extent should the defendant be held legally responsible for the consequences of that act? (Legal causation). The factual causation is more straightforward where the robot's misbehavior is the sole cause present [but for the robot to misbehave, the human would not be injured]. However, where multiple causes exist, proving causation under 'material contribution test' becomes difficult. A factory worker's injury resulted

¹² *Sinniah Nadaraja v. Ceylon Transport Board and Another* 1978 2 SLR 48, P. 53

¹³ *Kruger v. Coetzee* 1966 (2) SA 428.

from several contributing factors, including the decision of an AI-driven robot to mistakenly perceive the worker as an inanimate object, the worker's failure to follow established safety protocols, and the insufficient maintenance of the robot. It is also possible that it was the result of an intervening factor, such as data contamination (hacking) by a third party. Nevertheless, the AI-robot's decision is non-explainable as a result of the model's black-box nature. If the robot is substituted by a human, then he can be inquired and cross-examined thereby allowing the courts to investigate which contributing factor increased the risk.

On the other hand, proving causation-in-law is more difficult in the context of AI. As reiterated in this paper, the question 'what kind of damage is reasonably foreseeable' is difficult to ascertain in the context of AI due to unpredictability and multi-party involvement. For example, contemplate an autonomous delivery robot that harms a pedestrian as a result of a misclassification error by its artificial intelligence system. The damage may arise from AI's autonomous decision making, which developers or deployers may not have anticipated, particularly if it stemmed from infrequent edge situations or environmental variables. Conversely, the damage may be attributable to human stakeholders, including developers for insufficient training data, the logistics business for poor testing, or maintenance providers for sensor malfunctions. Moreover, third-party interference or the pedestrian's contributory fault could further confuse causality. There's a substantial possibility that each party may argue that the damage was not foreseeable from their respective perspectives. Opacity of the robot may further complicate finding the remoteness of the damage as the robot cannot explain the situation unlike a human being.

As previously noted, the primary challenge in establishing causation stems from the lack of transparency, or the so-called black box nature, of AI systems, particularly those developed using deep neural networks. However, during the previous years, much research has been conducted to achieve an explainable model to mitigate the black-box issue thereby, attempting to create trustworthy AI. As a result, explainable AI models (XAI) were introduced to the context the purpose of which was to clarify AI decision-making, its actions and its predictions. For instance, explainable models offer a post-hoc explanation through the use of LIME and SHAP methods. However, many doubt the accuracy and reliability of such explanations and the impact of transparency – accuracy tradeoff. Henry Frase et al using the ACCC v Trivago (AU) as an example demonstrated how the company deceived the courts through the use of XAI method in a case¹⁴.

Nonetheless, many authors consider XAI to be the most promising approach to establishing causation at this moment, despite its lack of accuracy. Consequently, the paper suggests a harmonized approach that includes the integration of post-hoc XAI explanations into the robot and the requirement for expert evidence to ensure accuracy. There are a number of XAI methods which provide post- hoc explanations from different perspectives. As a matter of fact, the courts

¹⁴ Henry Fraser, Aaron J. Snoswell, and Rhyle Simcock. 2022. AI Opacity and Explainability in Tort Litigation. In 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22), June 21-24, 2022, Seoul, Republic of Korea. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3531146.3533084>

do not mandate that the defendant establish legal and factual causation; rather, the judge is responsible for doing so by evaluating the facts and evidence of both parties. Therefore, if the model is capable of articulating its reasoning in a manner that is commensurate with that of a human defendant, it is sufficient to consider it as valid reasoning for its decision. As explained, the necessity for any AI entity to be endowed with the ability to record and emphasize on beginning and end of the process is of utmost importance¹⁵. An expert shall assess and evaluate evidence taken from the beginning and end of the process, thereby asserting the accuracy of the model's explanation.

4. Conclusion

In conclusion, the objective of this paper was to analyse the appropriateness of existing delictual liability regime of Sri Lanka in addressing physical injuries caused by AI-robots and to offer recommendations. It was argued that the inherent characteristics of AI comprising lack of explainability, unpredictability, autonomy and multi-party involvement clearly affect the attribution of Aquilian actions under delictual liability. It was further contended that the primary liability to be vested in developer, operator and end-user. Subsequently, pertaining to the Aquilian action, it's argued that while the infringement of legally recognized right can *prima facie* be established, the requirement whether it occurred in a legally reprehensible manner cannot be established due to the autonomy and unpredictability of AI-robot. Therefore, it was recommended that wrongfulness should be assessed in accordance with the nature of the harm and control of the risk. Conversely, it was argued that patrimonial loss could be established without much ambiguity in this context.

Referring to the *culpa*, the paper argued that the standard of an average reasonable person is untenable due to unpredictability and therefore, it's recommended that a professional standard shall be applied to human stakeholders. The paper argued that black-box nature of the DL methods utilized in AI-robot fails the conventional causation in-fact and in-law. The paper recommends that methods of explainable artificial intelligence be incorporated into the AI robot to generate a post hoc explanation, provided that expert evidence is available at both the beginning and end of the process to verify its accuracy, thereby enabling courts to infer causation in fact and in law. While the paper does not cover all potential issues in applying the Aquilian action to AI-robot cases, it addresses the key concerns and recommends implementation only after thoroughly assessing legal, technical, and practical considerations.

References

Primary Sources

¹⁵ Cheryl Seah, Liability arising from the use of Artificial Intelligence, Singapore Law Gazette, Available at: < Liability Arising from the Use of Artificial Intelligence - The Singapore Law Gazette>, Accessed: 8th March 2025.

1. Pelis And Another v. Arnaashal (1980) 2 Sri L.R.
2. Coronation Brick (Pty) Ltd v Strachan Construction Co (Pty) Ltd 1982 4 SA 371 (D) 380
3. Sinniah Nadaraja v. Ceylon Transport Board and Another 1978 2 SLR 48.
4. Kruger v. Coetzee 1966 (2) SA 428.

Secondary Sources

5. Barfield, Woodrow. "Liability for Autonomous and Artificially Intelligent Robots" Paladyn, Journal of Behavioral Robotics, vol. 9, no. 1, 2018, pp. 193-203. <https://doi.org/10.1515/pjbr-2018-0018>
6. Benhamou, Yaniv & Ferland, Justine. (2020). ARTIFICIAL INTELLIGENCE & DAMAGES: ASSESSING LIABILITY AND CALCULATING THE DAMAGES.
7. Berx N, Brescia A, Aqamarina R, Curcio EM, Pintelon L, Carbone G. Stakeholders' perspectives on safety related human–robot collaborative scenarios. International Journal of Advanced Robotic Systems. 2023;20(5). doi:10.1177/17298806231200095
8. Cheryl Seah, Liability arising from the use of Artificial Intelligence, Singapore Law Gazette, Available at: < Liability Arising from the Use of Artificial Intelligence - The Singapore Law Gazette>, Accessed: 8th March 2025.
9. Chinen MA. The co-evolution of autonomous machines and legal responsibility. Va J Law Technol. 2016;20(2):338-393.
10. EUROPEAN COMMISSION, Liability for emerging digital technologies, Accompanying the document Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee Regions, April 25, 2018, SWD/2018/137 final, available at: EUR-Lex - 52018SC0137 - EN - EUR-Lex
11. Henry Fraser, Aaron J. Snoswell, and Rhyle Simcock. 2022. AI Opacity and Explainability in Tort Litigation. In 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22), June 21–24, 2022, Seoul, Republic of Korea. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3531146.3533084>.
12. Lemley, Mark A. and Casey, Bryan, Remedies for Robots (July 31, 2018). 86 University of Chicago Law Review 1311 (2019), Stanford Law and Economics Olin Working Paper No. 523, Available at SSRN: <https://ssrn.com/abstract=3223621> or <http://dx.doi.org/10.2139/ssrn.3223621>.
13. Rachum-Twaig, Omri, Whose Robot Is It Anyway?: Liability for Artificial-Intelligence-Based Robots (February 21, 2019). University of Illinois Law Review, Vol. 2020, Forthcoming, Available at: SSRN: <https://ssrn.com/abstract=3339230> p.7-8.
14. Woodrow BARFIELD, "Liability for Autonomous and Artificially Intelligent Robots" in Paladyn Journal of Behavioral Robotics, 2018, Volume 9, Issue 1, pp. 193-203
15. Yampolskiy, Roman. (2020). Unpredictability of AI: On the Impossibility of Accurately Predicting All Actions of a Smarter Agent. Journal of Artificial Intelligence and Consciousness. 07. 109-118. 10.1142/S2705078520500034.