WHOSE ROBOT IS IT ANYWAY?: LIABILITY FOR ARTIFICIAL-INTELLIGENCE-BASED ROBOTS

*Omri Rachum-Twaig**

The idea of robots that possess autonomous capabilities and intelligence and are ungoverned by human directions or supervision dates as far back as several decades ago. These (then-) futuristic ideas were steadily incorporated into very real and current technology. Combined with Artificial Intelligence ("AI") technology, products and machines disrupt the idea of agency and the involvement of human beings in manufacturing and provision of services. How should liability be constructed when there is no apparent agency or personhood or when actions are almost inherently unforeseeable? More specifically, in the context of AI-based robots, do models of products liability or other tort liability fit the new framework? This Article seeks to explain why current law and doctrine, such as products liability and negligence, cannot provide an adequate framework for these technological advancements, mainly due to the lack of personhood, agency, and the inability to predict and explain robot behavior. Zooming out from specific doctrines, the Article also suggests that none of the three main liability regimes—strict liability, negligence, and no-fault mandatory insurance—adequately resolves the challenges of AI-based robots. Ultimately, this Article aims at suggesting supplementary rules that, together with existing liability models, could provide better legal structures that fit AI-based robots. Such supplementary rules will function as quasi-safe harbors or predetermined levels of care. Meeting them would shift the burden back to current tort doctrines. Failing to meet such rules would lead to liability. Such safe harbors may include a monitoring duty, built-in emergency brakes, and ongoing support and patching duties. The argument is these supplementary rules could be used as a basis for presumed negligence that complements the existing liability models. If adopted, they could establish clear rules or best practices that determine the scope of potential liability of designers, operators, and end-users of AI-based robots.

* PhD (Law), Tel Aviv University, Faculty of Law; Adjunct Professor, Tel Aviv University, Faculty of Law; Research Fellow, Federmann Cyber Security Research Center, The Hebrew University of Jerusalem, Faculty of Law. I wish to thank Mark Lelmey, Omer Yehezkel Pelled, Ohad Somech, and Asaf Wiener for reading earlier drafts of this Article and for their fantastic, thoughtful comments that tremendously helped me articulate my arguments. This work was supported by The Federmann Cyber Security Center in conjunction with the Israel National Cyber Directorate and by the UNIGE-HUJI Joint Seed Money Funding Scheme 2017-2018.

TABLE OF CONTENTS

I. INTRODUCTION

In March 2016, Microsoft shut down its artificial-intelligence-based chatbot, Tay, which had been developed to autonomously interact with users via Twitter and provide data for research on conversational understanding.[1] Tay was supposed to adapt itself and self-learn conversational skills by analyzing Twitter tweets.[2] The shut-down came following a series of racist and misogynic tweets by Tay, surprising both users and Microsoft.[3] A little over a year later, Facebook had to shut down its AI-based chatbot experiment.[4] After launching an experi-

1. Sarah Perez, *Microsoft Silences Its New A.I. Bot Tay, After Twitter Users Teach It Racism*, TECHCRUNCH (Mar. 24, 2016, 9:16 AM), https://techcrunch.com/2016/03/24/microsoft-silences-its-new-a-i-bot-tay-after-twitter-users-teach-it-racism; James Vincent, *Twitter Taught Microsoft's AI Chatbot to be a Racist Asshole in Less than a Day*, VERGE (Mar. 24, 2016, 6:43 AM), https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist.

2. *See* sources cited *supra* note 1.

3. *See* sources cited *supra* note 1.

4. Andrew Griffin, *Facebook's Artificial Intelligence Robots Shut Down After They Start Talking to Each Other in Their Own Language*, INDEPENDENT (July 31, 2017, 5:10 PM), https://www.independent.co.uk/life-

ment attempting to teach autonomous bargaining skills between chatbots, the Facebook developers noticed that the two bots, Alice and Bob, began interacting in an unintelligible manner. After not being able to decipher the codes used by the two bots, Facebook shut the project down.[5] These two bot cases had relatively harmless implications, but they suggest a potential pattern regarding the use and deployment of AI-based technology. What if, in the (near) future, Tay was able to establish substantial influence, leading to the actual formation of hate groups that inflict physical harm? In the near future—if this has not already occurred—robots could conspire to take over Facebook accounts and retrieve sensitive personal or financial information on individuals.

The idea of robots with autonomous capabilities and intelligence ungoverned by human directions or supervision dates back several decades.[6] These (then-) futuristic ideas were steadily incorporated into very real and current technology, first with respect to digital communications and cyberspace, and then, mainly in recent years, to the physical world as well.[7] In the age of connected devices and robotics, cyberspace is no longer limited to bits and bytes. Connected devices, personal-use machines, and robots allow activities in cyberspace to directly affect the physical world in a more concrete way than ever, not only concerning critical infrastructures, but also in our homes, at our workplaces, and on our roads.[8] Combined with AI technology, products and machines disrupt the idea of agency and the involvement of human beings in the provision of services and the manufacturing of consumer products. This trend and phenomenon is coupled with the unpredictable manner in which such robots behave and the inability to foresee risks that they may inflict.[9]

---

style/gadgets-and-tech/news/facebook-artificial-intelligence-ai-chatbot-new-language-research-openai-google-a7869706.html.

5. *Id*. The actual negotiations between the bots proceeded as follows:

Bob: i can i i everything else . . . . . .
Alice: balls have zero to me to me to me to me to me to me to me to me to
Bob: you i everything else . . . . . .
Alice: balls have a ball to me to me to me to me to me to me to me
Bob: i i can i i i everything else . . . . . .
Alice: balls have a ball to me to me to me to me to me to me to me
Bob: i . . . . . .
Alice: balls have zero to me to me to me to me to me to me to me to me to
Bob: you i i i i i everything else . . . . . .
Alice: balls have 0 to me to me to me to me to me to me to me to me to
Bob: you i i i everything else . . . . . .
Alice: balls have zero to me to me to me to me to me to me to me to me to.

6. *See*, *e.g.*, ISAAC ASIMOV, I, ROBOT (1950); ARTHUR C. CLARKE, 2001: A SPACE ODYSSEY (1968); *see also* David C. Vladeck, *Machines Without Principals: Liability Rules and Artificial Intelligence*, 89 WASH. L. REV. 117, 117 (2014) (arguing that the idea of self-thinking machines dates back to the dawn of civilization).

7. Andrew Meola, *What Is the Internet of Things? What IoT Means and How it Works*, BUS. INSIDER (May 10, 2018, 12:06 PM), https://www.businessinsider.com/internet-of-things-definition.

8. *Id*.

9. Karl Frederick Rauscher, *Can We Avoid the Potential Dangers of AI, Robots and Big Tech* Companies?, SCI. AMERICAN (Nov. 26, 2018), https://blogs.scientificamerican.com/observations/can-we-avoid-the-potential-dangers-of-ai-robots-and-big-tech-companies/.

In light of these technological advancements, important questions relating to liability arise. How should liability be constructed in the absence of any apparent agency or personhood, or when actions are almost inherently unforeseeable? More specifically, in the context of AI-based robots, do models of product liability or other tort liability fit the new framework? Should designers of AI-based robots be strictly liable for damages inflicted by their creations? Should programmers of autonomous robots be liable for all the robots' expected and unexpected future conduct and actions? Current forms of liability seem to be insufficient to capture the entire spectrum of possibilities and nuances that arise in the context of AI-based robots.

This Article seeks to explain why current law and doctrine cannot provide an adequate framework for these technological advancements. It begins by mapping specific features of AI-based robots such as the new or augmented types of harms that may be caused, the lack of personhood and agency, and the impossibility of foreseeing and explaining certain behaviors the robots exhibit.

It will then review dominant tort doctrines to expose their shortcomings in the context of AI-based robots. For example, products liability doctrines (as well as other tort doctrines) are commonly restricted to physical injuries and damage to property and cannot necessarily account for other types of damages such as privacy violations, pure economic harm, denial of critical services, and the like.[10] Moreover, they are generally limited to harm caused by design and manufacturing defects, a concept that does not easily fit the idea of AI. In addition, other general forms of liability in torts are not adequate for several reasons. Tort law generally requires agency as a precondition.[11] In the age of AI and autonomous machines, however, the question of agency may pose challenges to which, in the absence of the legal accountability of robots, tort law cannot necessarily respond. Negligence is also insufficient because the duty of care and the standards for reasonable precautions depend on a baseline that is constantly changing in these technological fields and is disrupted by new types of unexpected harms and a general lack of foreseeability, which undermine the concept of breach of duty and the general concept of causation.

Zooming out from specific doctrines, the Article will review three general tort liability regimes and discuss whether any of them fit the AI challenges. It will suggest that strict liability regimes may impose an excessive burden on persons utilizing AI-based robots since the ultimate purpose of such products is to function in an unpredictable manner that the manufacturer cannot necessarily foresee. Thus, manufacturers are not necessarily better situated to assess the risks and the ways to prevent them. Negligence as a liability regime appears to be inadequate as well due to the expected difficulty of courts to set the optimal level of care in the context of AI. Even full no-fault mandatory insurance schemes cannot necessarily overcome these shortcomings, because of the difficulty in setting premiums and assessing potential risk as well as the cross-jurisdictional nature of AI-robotics.

10. *See* RESTATEMENT (THIRD) OF TORTS: PRODS. LIAB. § 1 (AM. LAW INST. 1998).
11. RESTATEMENT (THIRD) OF AGENCY § 7.03 (AM. LAW INST. 2006).

Ultimately, this Article aims to suggest supplementary rules that, together with existing liability models, could provide better legal structures that fit these business and technological requirements, at least for the near future and in the absence of the legal liability of robots. Such supplementary rules would function as quasi-safe-harbors or predetermined levels of care. Meeting them would grant immunity from specific doctrines, such as products liability, and would shift the burden of proving negligence back to potential plaintiffs. Failing to adhere to such rules would lead to liability. Such supplementary rules may include a monitoring duty, built-in emergency brakes, and ongoing support and patching duties. The argument is that these supplementary rules could be used as a basis for presumed negligence that complements the existing liability models. If adopted, they could establish clear rules or best practices that determine the scope of potential liability of designers, operators, and end-users of AI-based robots. Such models of presumed negligence and quasi-safe harbors may fit those circumstances in which harms caused by AI-based robots disrupt current tort doctrines. Naturally, AI-based robots will function in various ways, some of which may not raise such difficulties. Thus, different liability models would apply to different phenomena associated with AI-based robots.

II. THE CHALLENGES OF AI-BASED ROBOTS

Before we delve into the specific challenges that AI-based products and machines pose to the legal thought, we must first try to articulate the terms that will be used throughout this Article. There is substantial literature attempting to define and articulate the features of self-operating devices and machines, usually referred to as robots. As Ryan Calo put it in 2015, "robots are best thought of as artificial objects or systems that sense, process, and act upon the world to at least some degree."[12] He bases this definition on the technological "sense-think-act" paradigm attempting to define the technological concept of robotics.[13] Suggesting a definition for robotics, Calo emphasizes three essential qualities of robots–embodiment, emergence, and social valence.[14] In Calo's words:

> Robotics combines, arguably for the first time, the promiscuity of information with the [embodied] capacity to do physical harm. Robots display increasingly emergent behavior, permitting the technology to accomplish both useful and unfortunate tasks in unexpected ways. And robots, more so than any technology in history, feel to us like social actors.[15]

While the three essential features of robotics suggested by Calo are important, not all of them are key for the purposes of this Article. The digital-physical interface, while important and unique in the context of cyberlaw, does not necessarily yield unique results in the context of tort liability. Moreover, AI-

12. Ryan Calo, *Robotics and the Lessons of Cyberlaw*, 103 CALIF. L. REV. 513, 531 (2015).

13. *Id*. at 529; *see also* ROLF PFEIFER & CHRISTIAN SCHEIER, UNDERSTANDING INTELLIGENCE 37 (1999); Rodney A. Brooks, *Intelligence Without Reason*, 1 PROC. 12TH INT'L JOINT CONF. ON ARTIFICIAL INTELLIGENCE 569, 570 (1991).

14. Calo, *supra* note 12, at 532.

15. *Id.* at 515.

based robots may sometimes seem completely disembodied and yet still raise important questions and challenges to existing legal frameworks.[16] In this context, I follow Mark Lemley and Bryan Casey who suggest that the definition of robots includes both physical embodiments referred to by Calo, as well as strictly digital phenomena that follow the same lines.[17]

Lemley and Casey are right to observe that the terms AI and robotics are commonly used interchangeably, and therefore they propose a working definition that includes "any hardware or software system exhibiting intelligent behavior."[18] For our purposes, I offer to deconstruct such a definition and keep exploring one of its components. I suggest that the first component—robots, machines, or hardware and software—refers to the nonhuman agent capable of demonstrating AI. This is important for two reasons. First, it creates the distinction between the acts of humans (coders or designers for example) from the acts of the robots they create. Second, it inevitably emphasizes the more important factor for the purpose of this Article: AI.

The concept of AI has existed for quite a while.[19] In its first wave, also known as good old-fashioned AI ("GOFAI"), it was an extensive use of common algorithms and code that instructed machines to make specific decisions and act in specific circumstances.[20] The level of specificity of such algorithms led to sophisticated machines capable of performing tasks as if they were intelligent on their own account, but in fact, any decision made by the machine could be traced back to the instructions given by the designer.[21] GOFAI faced a challenge: how to scale the capabilities of machines to overcome edge-scenarios and achieve their goals in indeterminate situations.[22] The introduction of machine-learning capabilities overcame this challenge.[23]

The basic idea of machine-learning is instead of tackling all possible scenarios and setting specific instructions for each of them, the designers set a goal for the machine, and by complex stages of repeated experiments and self-research, the program eventually writes its own optimized instruction to reach the

16. See, for example, the discussion on privacy harms *infra* note 41, and the discussion on as-a-service robots, *infra* notes 166–68.

17. Mark A. Lemley & Bryan Casey, *Remedies for Robots*, 86 U. CHI. L. REV. 1311, 1319–21 (2019).

18. *Id.*

19. Scholarship on the concept of AI dates back at least to the late 1950s. *See, e.g.*, J. MCCARTHY ET AL., ARTIFICIAL INTELLIGENCE RESEARCH LABORATORY OF ELECTRONICS AT THE MASSACHUSETTS INSTITUTE OF TECHNOLOGY (1959); MARVIN L. MINSKY, SOME METHODS OF ARTIFICIAL INTELLIGENCE AND HEURISTIC PROGRAMMING 5 (1958) (exploring "ideas concerning the design or programming of machines to work on problems for which the designer does not have, in advance, practical methods of solution"). A journal dedicated to AI was first published in 1970. *See* 1 ARTIFICIAL INTEL. (1970). Publication of legal thoughts on AI started in the early 1990s. *See, e.g.*, Lawrence B. Solum, *Legal Personhood for Artificial Intelligences*, 70 N.C. L. REV. 1231 (1992).

20. Lemley & Casey, *supra* note 17, at 1323.

21. *Id.* at 1322.

22. *Id*. at 1324.

23. *Id.* at 1324–26.

predetermined goal.[24] This is a reflection of the common saying that the "program becomes the programmer."[25] In fact, there is much to it. Machine-learning capabilities lead to scenarios in which the program or machine reaches a predetermined goal, but its programmers may not have an exact understanding of how it reached such goal or what the stages leading to success were.[26] This is exactly the starting point of the working definition of AI for the purpose of this Article. In other words, the focus here would be on the characteristics of AI that lead to actions that are either unexplainable or unforeseeable to the robots' designers (or human beings in general) but that is nonetheless an inherent feature of the technology.[27]

This conceptualization of AI could obviously vary in degree and apply differently to distinctive circumstances. Formal technological education is not required to assume that the more precise and well-defined the goal or problem is (and the more predictable the operation environment is and the simpler the case is), the more predictable and explainable the robot's behavior will be. But when we move forward to more complex goals and problems, perhaps sometimes purposefully ill-defined,[28] in sophisticated environments requiring multi-machine-human interaction, the robot's behavior will be less stable, predictable, and explainable.[29] This could also be reflected by the technical methods employed in the context of machine-learning AI. One main machine-learning method—supervised learning—may yield more predictable results. Supervised learning refers to a method in which the data used for the training and learning process of the program are prelabeled by the designers, thus having a great effect on the learning process.[30] Another significant method—unsupervised learning—may (as evident from its name) yield less predictable results, as it allows the program to learn based on unlabeled data that the program labels autonomously.[31] Yet

24. *Id.*

25. *See, e.g.*, Kaushik Chatterjee, *Unearthing the Layers of Machine Learning*, MEDIUM (Sept. 17, 2019), https://medium.com/@kchatr/unearthing-the-layers-of-machine-learning-20b2738758ea.

26. Lemley & Casey, *supra* note 17, at 1336. For a concise but more technical general description of machine-learning, see M. I. Jordan & T. M. Mitchell, *Machine Learning: Trends, Perspectives, and Prospects*, 349 SCI. 255 (2015).

27. This is obviously dependent on how foreseeability is interpreted for various purposes. This will be discussed *infra* Part II.C

28. The concept of ill-defined problems is often used in the field of creativity studies to differentiate between standard problem-solving procedures and more open-ended creative activity. *See* OMRI RACHUM-TWAIG, COPYRIGHT LAW AND DERIVATIVE WORKS: REGULATING CREATIVITY 29–31 (2019); Omri Rachum-Twaig, *Recreating Copyright: The Cognitive Process of Creation and Copyright Law*, 27 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 287, 310–13 (2016). It could well be expected that when ill-defined goals or problems are fed to a machine-learning process, the results of the AI-based robot will be highly unpredictable and perhaps creative. For a legal discussion on AI-based "creativity" see, for example, Annemarie Bridy, *Coding Creativity: Copyright and the Artificially Intelligent Author*, 2012 STAN. TECH. L. REV. 5 (2012); Ralph D. Clifford, *Intellectual Property in the Era of the Creative Computer Program: Will the True Creator Please Stand Up?*, 71 TUL. L. REV. 1675 (1997); Robert C. Denicola, *Ex Machina: Copyright Protection for Computer-Generated Works*, 69 RUTGERS U. L. REV. 251 (2016); Robert Yu, *The Machine Author: What Level of Copyright Protection is Appropriate for Fully Independent Computer-Generated Works*, 165 U. PA. L. REV. 1245 (2017).

29. Lemley & Casey, *supra* note 17, at 1334.

30. Jordan & Mitchell, *supra* note 26, at 257–58.

31. *Id.* at 258.

another trend in machine-learning may be even more unstable. New research on machine-learning attempts to develop a method in which the program learns various skills, and it keeps learning new skills based on that underlying skillset.[32] This creates an everlasting process of learning, much like human learning and development skills.[33]

The purpose of this preliminary note is not to account for all current or future trends of AI. Rather, it is intended to clarify my meaning in referring to AI throughout this Article. In addition, it must be noted that not any application or use of AI will necessarily raise the legal concerns discussed below. There may well be cases in which the use of AI will be part of a human-supervised process and could be encompassed by the current legal doctrines. In this sense, as Karny Chagal-Feferkorn put it, AI "is a spectrum."[34] The purpose of this Article is to account for that part of the spectrum that inherently challenges our current legal doctrines and regimes, as will be discussed below; the use of AI-based robots that fully meet the criteria set forth above and below falls within this range.[35]

To summarize this Part, while the term *robot* will refer to nonhuman agents capable of demonstrating AI, the term AI will refer to the program on which the robot runs and that causes it to act in a manner that is, inherently, either unexplainable or unforeseeable to humans. This may not accurately describe contemporary AI-based robots or may account only for parts of current technology.[36] Autonomous vehicles, for example, are enhanced with significant AI capabilities.[37] But since they are designed to operate in an environment that is constrained by strict rules (physical and man-made) and achieve very specific goals, at least some behavior of such robots may be largely predictable or at least explainable in retrospect.[38] I believe, however, that the principles discussed in this Article, as well as the normative suggestions could apply to future technology

32. *Id.* at 259–60.

33. *Id.* at 260.

34. *See* Karni A. Chagal-Feferkorn*, Am I an Algorithm or a Product? When Products Liability Should Apply to Algorithmic Decision-Makers*, 30 STAN. L. & POL'Y REV. 61, 72–73 (2019).

35. For a suggestion on how to determine whether the deployment of a specific form of AI should be considered merely a part of a products under current product liability doctrine, or a special category of AI-based or autonomous robots, see *id.*

36. This is true, for example, for AI-based robots that account for only part of the process or goal expected to be achieved, while the rest of the process is handled by human beings, or to robots based on simpler versions of AI such as GOFAI. Lemley and Casey cite Jonathan Zittrain as describing a phenomenon of "autonomish" robots. Lemley & Casey, *supra* note 17, at 1318 n.19 (citing Jonathan L. Zittrain, *What Yesterday's Copyright Wars Teach Us about Today's Issues in AI*, delivered as the David L. Lange Lecture in Intellectual Property Law at Harvard Law School (2018), transcript archived at http://perma.cc/TZP7-H4EH).

37. Mark A. Geistfeld, *A Roadmap for Autonomous Vehicles: State Tort Liability, Automobile Insurance, and Federal Safety Regulation*, 105 CALIF. L. REV. 1611, 1644–47 (2017).

38. As will be discussed below, the main literature on the legal challenges of robots and AI revolves around autonomous vehicles. *See, e.g.*, *id.*; Kenneth S. Abraham & Robert L. Rabin, *Automated Vehicles and Manufacturer Responsibility for Accidents: A New Legal Regime for a New Era*, 105 VA. L. REV. 127 (2019); Gary E. Marchant & Rachel A. Lindor, *The Coming Collision Between Autonomous Vehicles and the Liability System*, 52 SANTA CLARA L. REV. 1321 (2012); Bryant Walker Smith, *Automated Driving and Product Liability*, 2017 MICH. ST. L. REV. 1, 74 (2017). I do not believe that autonomous vehicles are a representative case study to review the challenges of AI, at least as defined in this Article. It is perhaps due to this that this Article reaches different doctrinal and normative conclusions about AI in comparison to some of the above-mentioned literature.

that fully meets the AI definition used here, as well as to some circumstances that arise out of the use of contemporary robots.

A. Augmented Harms

In the history of product-related harms, the most obvious cases tend to be those of physical injury, damage to property, and the loss of the product itself due to malfunction.[39] But in the context of AI-based robots, while preserving significant presence in the physical realm,[40] the array of potential harms widens, as additions to the product include a new facet: intelligence. In fact, the product ceases to merely be a passive physical (or digital) object and becomes, to a great extent, an active subject. Moreover, if the product or machine is connected, it may have a much more direct effect on other connected products or machines and even humans who participate in the network.

This new characteristic of AI-based products may lead to new or augmented types of harms that the law in general, and tort law and product liability law in particular, has not yet encountered in the context of products, or it at least may significantly augment existing harms. An immediate example could be privacy-related harms. Imagine that your AI-based personal-assistant bot, to which you eventually disclose a significant amount of sensitive personal data, such as your medical status, your financial status, and your personal affairs, decides to disclose this information to a third party without your consent.[41]

Another type of harm that should be considered in the context of AI-based products is autonomy-related harm such as loss of autonomy. Imagine an AI-based bodyguard or home-safety robot that manages entry permissions to one's home. Circumstances could lead the robot to believe that one of the family members is a burglar thus locking the family member in the house and potentially leading to circumstances of false imprisonment.[42] This could be extended to broader concepts of mere harms to autonomy that pose a challenge to tort law at large but may become more frequent and present in the context of AI-based robots.[43]

39. RESTATEMENT (THIRD) OF TORTS: PRODS. LIAB. § 1 (AM. LAW INST. 1998) ("The rule stated in this Section applies only to harm to persons or property, commonly referred to as personal injury and property damage.")

40. *See* Calo, *supra* note 12, at 513.

41. For the purpose of this example, assume that it does not disclose the data to its manufacturer, and that it may genuinely believe that the disclosure of the data could benefit the user. This is to differentiate this example from more traditional cases of personal data collection and disclosure by 'always-on' connected devices and personal assistants such as the Amazon Echo. *See* Eldar Haber, *Toying with Privacy: Regulating the Internet of Toys*, 80 OHIO ST. L. REV. 399, 407 (2019).

42. This could also be the result of other types of interactions with the robot or machine, and perhaps even a "positive" behavior on the part of the robot attempting to prevent what it believes to be an unauthorized attempt to enter the house. A similar result already occurred in the context of cybersecurity breaches to smart-home systems, deploying a ransomware that effectively locked the guests of an Austrian hotel in their rooms. *See* Dan Bilefsky, *Hackers Use New Tactic at Austrian Hotel: Locking the Doors*, N.Y. TIMES (Jan. 30, 2017), https://www.nytimes.com/2017/01/30/world/europe/hotel-austria-bitcoin-ransom.html.

43. An example could be discriminatory conduct of AI-based robots that is not driven by a discriminatory design. *See, e.g.*, Jeffrey Dastin, *Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women*,

Finally, AI-based robots could potentially cause pure economic harm in a variety of cases. Tort law is generally reluctant to afford damages for pure economic harms and limits itself to damage to property and personal injuries.[44] Since AI-based robots are expected to be integrated into everyday human activities, including financial activities, it would come as no surprise if the conduct of such robots led to economic losses that tort law struggles to cover.[45]

B. (The Lack of) Agency and Personhood

Perhaps the most basic concept in legal liability in general and tort liability, in particular, is that the law governs the behavior of people and liability could only be attributed to a person demonstrating the capability to act as a purposive agent.[46] This is clearly manifested by the Restatement (Second) of Torts with respect to liability in tort.[47] It instructs us that, "[t]he word 'actor' is used throughout the Restatement of this Subject to designate either the person whose conduct is in question as subjecting him to liability toward another, or as precluding him from recovering against another whose tortious conduct is a legal cause."[48] As the liability for most intentional torts, as well as negligence, is defined as the liability of actors for their acts or conduct, tort law generally holds persons directly liable only for wrongful acts or omissions.[49]

But AI-based robots are not human. Therefore, at least as of now and in the near future, it seems almost impossible to attribute any concept of personhood or agency to such robots, at least legally. If this is the case, the immediate result is that AI-based robots cannot currently be directly liable for their actions and conduct, even those that inflict harm.[50] The law, however, does not necessarily stop there. Several legal concepts extend the liability of individuals beyond their immediate acts. One example is the doctrine of product liability that could potentially apply to AI products in the same manner as it generally applies to other products.[51] This will be discussed at greater length below. Another broader example is the direct or vicarious liability of the principal for the actions of an agent in various circumstances.[52] Here too, the law assumes that both principal and agent are persons capable of pursuing their free will, where the latter performs

REUTERS (Oct. 9, 2018, 10:12 PM), https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G.

44. *See* Geistfeld, *supra* note 37, at 1630.

45. *See id.*

46. The term 'agent' here does not refer to the legal concept of the principal-agent relationship (although this concept will be discussed soon), rather the concept of an actor demonstrating free will and able to control her actions.

47. RESTATEMENT (SECOND) OF TORTS § 3 (AM. LAW INST. 1965).

48. *Id.*

49. This concept was obviously extended to apply to legal persons such as corporations as well under various doctrines. For a historical review of the concept of corporation as legal persons see John Dewey, *The Historic Background of Corporate Legal Personality*, 35 YALE L.J. 655 (1926).

50. *See generally* Vladeck, *supra* note 6.

51. *See infra* note 72.

52. *See generally* RESTATEMENT (THIRD) OF AGENCY §§ 7.03–7.07 (AM. LAW INST. 2006).

actions on behalf of the former.[53] But even if we are able to take a step further and assume that the concept of principal liability could apply to AI-based products and that they could be seen as agents, significant difficulties remain.

The first question would be determining whether a principal-agent relationship had been established. Such a relationship is created "by a principal's manifestation to an agent that, as reasonably understood by the agent, expresses the principal's assent that the agent take action on the principal's behalf."[54] Applying this to the context of AI-based robots is a struggle. Different stakeholders could be considered as principals in this type of relationship. When a corporation (whether designing the product or distributing it) is actually operating it, we may think of such corporation as operating the robot on its behalf.[55] In other cases, a user may be considered as a principal with respect to a machine that it operates, while the designer of such a robot would likely not be considered a principal in this context.

The second question would be determining whether liability could be attributed to the principal for the machine's conduct as an agent. The Restatement provides two general concepts for such liability: direct and vicarious.[56] For direct liability to apply, it must be established that the agent's acts are tortious, that if they had been performed by the principal they would have been considered tortious, or that the principal was negligent in selecting, supervising, or controlling the agent.[57] For vicarious liability to apply, it must be established that the agent is an employee of the principal, acting within the scope of employment.[58] It is questionable whether a robot can be considered an employee. While we can imagine circumstances in which robots replace people as service providers, thus effectively acting as employees, there surely may be various circumstances in which robots cannot be even remotely considered as employees. Even if we consider this concept a valid option, the Restatement defines an employee for these purposes as "an agent whose principal controls or has the right to control the manner and means of the agent's performance of work."[59] When considering AI-

53. *Id.* § 1.01 ("Agency is the fiduciary relationship that arises when one person (a 'principal') manifests assent to another person (an 'agent') that the agent shall act on the principal's behalf and subject to the principal's control, and the agent manifests assent or otherwise consents so to act."). The foregoing terminology clearly indicates that this type of relationship requires two persons capable of making decisions, including legal bodies, such as corporations and governments, which are ultimately constituted of humans. *See id.* § 1.04 (extending the term "person" to any organization that is capable of possessing rights and incurring obligations).

54. *Id.* § 3.01. It should be noted that all terms defined in the Restatement with respect to the formation of such relationship and the concept of actual authority are very human-centric, in the sense that they address the manifestation of the authority of the principal over the agent and the understanding of the agent of such manifestation. "Manifestation," however, is defined in § 1.03 to include not only spoken or written words, rather also any other conduct. *Id.* § 1.03. In this sense, it may be possible to extend the concept of agency and actual authority to apply to a relationship between a person and a machine, to the extent that such machine is capable of understanding (which is exactly our case).

55. RESTATEMENT (THIRD) OF AGENCY § 1.04 (AM. LAW INST. 2006) (extending the term "person" to any organization that is capable of possessing rights and incurring obligations).

56. *Id.* § 7.03.

57. *Id.* §§ 7.03(1)(a)–(b), 7.04, 7.05.

58. *Id.* §§ 7.03(2)(a), 7.07.

59. *Id.* § 7.07(3).

based robots, this becomes a challenge. When do we assume that the principal has control over the robot? If we believe that some acts of the robot are based on AI that is unpredictable and perhaps unexplainable by the principal, can we reasonably portray this dynamic as control over the robot's performance?

To be clear, I am not suggesting that the acts of AI-based robots could never be attributed to a human being. This could obviously apply in cases of proper AI-based robots and cases of "autonomish" robots. For example, if a service provider utilizes an AI-based robot to provide a specific service, say, serving beverages at a diner, such service provider may well be considered a principal of such robot with respect to harms caused to diners in their relationship with the robot during a meal.[60] My argument is that the basic technological characteristics of AI, however, will inevitably lead to many more circumstances in which the personhood-agency factor will simply not apply because no human being could be considered the principal behind the AI-robot acts. It is in these cases, I argue, that current tort doctrines and basic tort principles will fall short.

C. Unforeseeable Outcomes by Definition

A central pillar of tort law liability is the question of foreseeability.[61] This makes sense as we would usually like people to think about the consequences of their prospective actions before they actually engage in them, and we limit such demand to whatever is reasonably foreseeable (as we are only humans, after all).[62] But again, AI-based robots are not human. Not only that, but they are also by definition designed to act in unforeseeable ways. This is simply so because if we were able to predict all possible choices an automated product would have to make and predetermine such choices during the design stage, it would not be an AI-based robot at all, and AI would have been redundant.[63]

One could argue that we can make AI foreseeable, for example by embedding ground rules that will bypass any autonomous decision-making of the robot.[64] This is perhaps possible to a certain extent. For example, we may be able to include certain basic principles as ground rules for an AI-based robot, such as

60. *Id*.

61. The foreseeability question is raised in various tests of the applicability of certain tort doctrines, such as causation (in assessing a defendant's proximate cause) across all tort doctrines as well as the duty itself under negligence for example. *See, e.g.*, Omri Ben-Shahar, *Causation and Foreseeability*, *in* TORT LAW AND ECONOMICS § 3.12 (Michael Faure, ed., 2009); Guido Calabresi, *Concerning Cause and the Law of Torts: An Essay for Harry Kalven, Jr.*, 43 U. CHI. L. REV. 69 (1975); Jonathan Cardi, *Reconstructing Foreseeability*, 46 B.C. L. REV. 921 (2005) (arguing against the use of foreseeability to determine the duty in negligence); Leon Green, *Foreseeability in Negligence Law*, 61 COLUM. L. REV. 1401 (1961); William M. Landes & Richard A. Posner, *Causation in Tort Law: An Economic Approach*, 12 J. LEGAL STUD. 109 (1983); Richard W. Wright, *Causation in Tort Law*, 73 CALIF. L. REV. 1735 (1985); Benjamin C. Zipursky, *Foreseeability in Breach, Duty, and Proximate Cause*, 44 WAKE FOREST L. REV. 1247 (2009).

62. *See* Green, *supra* note 61, at 1411 ("It is not a matter of probabilities, more or less, but whether a prudent person would consider such a risk reasonable.").

63. For a similar general assertion with respect to robotics, see Calo, *supra* note 12, at 554–55.

64. *See, e.g.*, Bryan L. Casey, *Amoral Machines, or: How Roboticists Can Learn to Stop Worrying and Love the Law*, 111 NW. U. L. REV. 1347, 1347–66 (2017).

a do-not-kill rule.[65] But there is a limit to the amount and effect of rules that could be included as bypass rules for AI. Embedding enough rules to make all acts of a robot foreseeable necessarily undermines the basic principles of machine learning as explained above.[66] Thus, to the extent that we accept the desirability of AI-based robots, we must also accept a significant degree of unpredictability.

To be clear, I do not argue that any act of an AI-based robot would be unforeseeable. There are many circumstances in which AI-based robots could inflict foreseeable harms, much like harms resulting today from similar activities. For example, it seems foreseeable for an autonomous vehicle to improperly cross at a red light for an unexplainable reason and cause physical injuries and damage to property. In such circumstances, the fact that the exact actions leading to the harms are unexplainable and perhaps unforeseeable to begin with does not lead to the conclusion that the harms themselves are unforeseeable.[67] This could also be framed as the question of whether the *specific* harm or the *type* of harm is foreseeable.

We generally accept that foreseeing a type of harm would be enough to pass the foreseeability threshold in tort law, as would potentially be the case with autonomous vehicles. But I argue that, when AI-based robots reach their technological peak, there will not only be many cases in which both the actions themselves and the specific harms caused are unforeseeable, but also many cases in which the types of harms inflicted by the AI are unforeseeable. My point is that machine-learning capabilities defy the extensive concept of foreseeability of types of harms integral to current tort law doctrines. This is because machine-learning, when embedded in AI-based robots, may ultimately lead to such robots being capable of causing almost any type of harm. If we accept a broad concept of foreseeability, it may well be that harms would be considered foreseeable with respect to AI-based robots, undermining the important role of foreseeability in tort law. Thus, I argue that AI-based robots raise challenging questions with respect to this main element of tort law doctrines. In other words, as foreseeability is not merely a technical concept, rather it is also (and perhaps more so) a normative concept, I argue that determining the proper normative standard of foreseeability, taking into account the technical difficulties raised by AI, poses a significant challenge to current tort doctrines.[68]

65. *See, e.g.*, Sundar Pichai, *AI at Google: Our Principles*, GOOGLE (June 7, 2018), https://www.blog.google/technology/ai/ai-principles/ (detailing Google's recently published basic rules for AI). Google's principles are very general and could hardly be characterized as rules that will make the acts of AI always foreseeable.

66. *See infra* Part III.

67. *See* Calo, *supra* note 12, at 555.

68. For a demonstration of the normative aspects of foreseeability, see George G. Triantis, *Contractual Allocations of Unknown Risks: A Critique of the Doctrine of Commercial Impracticability*, 42 U. TORONTO L.J. 450, 465 (1992). Triantis suggests four levels of abstraction for events that could be considered foreseeable in the context of potential market-cost increases. *Id*. Choosing the right level of abstraction is both a technical and a normative question.

This leads to a significant challenge to the law in general and tort law in particular. Since AI-based products are, to a great extent, unforeseeable by design,[69] many tort law doctrines may not apply to the related human beings due to lack of foreseeability.[70] The alternative would be to determine that unforeseeable actions of AI-based products are always foreseeable, thus imposing liability for many potential risks and harms that could not be avoided by the liable tortfeasor.

III. THE SHORTCOMINGS OF EXISTING TORT LIABILITY MODELS

Now that we have a working definition of AI-based robots and after reviewing their inherent challenges and special features, we can proceed with reviewing whether existing tort liability models are prepared to tackle these challenges. In this Part, I will argue that all major tort doctrines that account for interpersonal liability and specifically for liability in the consumer consumption context cannot adequately govern the actions of AI-based robots and allocate among the relevant stakeholders the risks associated with the robots' actions.

A. Products Liability

Products liability law seems to be the most adequate arena for discussing liability for AI-based robots. After all, as intelligent as they may be, in many cases these are (still) products manufactured, distributed, and sold to consumers.[71] This is exactly the subject matter of products liability law, as set forth in the Restatement (Third) of Torts: Products Liability: "One engaged in the business of selling or otherwise distribut[ing] products who sells or distributes a defective product is subject to liability for harm to persons or property caused by the defect."[72] But AI-based robots pose significant challenges to current products liability doctrine that could potentially undermine its applicability in this context.[73]

69. For a similar observation see Lemley & Casey, *supra* note 17, at 1334–38 (referring to this phenomenon as unforeseen harms).

70. This includes product liability, negligence, and strict liability doctrines as will be elaborated in greater depth *infra* Part III.

71. Stephen Hennessey, *7 New Ways Golf Instruction is Embracing Artificial Intelligence and Innovative Technology*, GOLFDIGEST (Jan. 23, 2020), https://www.golfdigest.com/story/7-new-ways-golf-instruction-is-embracing-artificial-intelligence-and-innovative-technology.

72. RESTATEMENT (THIRD) OF TORTS: PRODS. LIAB. § 1 (AM. LAW INST. 1998). It is no surprise that many commentators on liability in the context of AI have focused on the product liability doctrine. *See, e.g.*, Geistfeld, *supra* note 37, at 1632; Gary E. Marchant & Rachel A. Lindor, *The Coming Collision Between Autonomous Vehicles and the Liability System*, 52 SANTA CLARA L. REV. 1321, 1322–24 (2012); Vladeck, *supra* note 6, at 130.

73. David Vladeck suggests that the product liability doctrine could easily apply to AI-based products to the extent that we can conceive of such products as agents of a human being. As explained below, I believe that this assertion is questionable. At least as far as fully AI-based products are concerned, however, and insofar as such products behave in a manner unrelated to explicit instructions by human beings, I believe that Vladeck could agree with the discussion below. See Vladeck, *supra* note 6, at 150.

Products liability doctrine mainly revolves around three triggers of liability: manufacturing defects,[74] design defects,[75] and failure to duly instruct or warn consumers.[76] The doctrine regarding manufacturing defects finds that a manufacturer will be liable for harms caused by a product with an unintentional defect, counter to the intended manufacturing specifications.[77] This trigger of liability does not raise unique concerns in the context of AI-based robots. If an AI-based robot does not function as intended due to a manufacturing failure or defect against the manufacturer's specifications, the products liability doctrine could easily apply as with any other "dumb" product.[78] The challenges in the AI context precisely arise where the product does function as intended. These cases are covered by the other basic triggers of the doctrine.

Under the second trigger, a manufacturer will be liable for harms caused by a product:

> when the foreseeable risks of harm posed by the product could have been reduced or avoided by the adoption of a reasonable alternative design by the seller or other distributor, or a predecessor in the commercial chain of distribution, and the omission, of the alternative design renders the product not reasonably safe.[79]

There are generally two approaches to determine whether a design defect has occurred.[80] The explicit approach provided by the Restatement is the "risk-utility" test, meaning that the plaintiff must prove that an alternative design could have

74. RESTATEMENT (THIRD) OF TORTS: PRODS. LIAB. § 2(a) (AM. LAW INST. 1998). For an elaborate review, see generally David G. Owen, *Manufacturing Defects*, 53 S.C. L. REV. 851 (2002).

75. RESTATEMENT (THIRD) OF TORTS: PRODS. LIAB. § 2(b) (AM. LAW INST. 1998). For an elaborate review, see generally David G. Owen, *Design Defects*, 73 MO. L. REV. 291 (2008).

76. RESTATEMENT (THIRD) OF TORTS: PRODS. LIAB. § 2(c) (AM. LAW INST. 1998).

77. *Id.* § 2(a).

78. For an example of how this trigger could apply in the case of autonomous cars, see Vladeck, *supra* note 6, at 140–41. The manufacturing defect trigger is also less likely to apply to the AI part of AI-based products because it applies only to units that depart from the intended design during manufacturing. In the context of software and algorithms, which are the core of AI, this is less likely to happen. While software bugs and failed design may well exist, these will be reviewed under the design defect part of the doctrine, as they will be consistent throughout all manufactured units. *See* Geistfeld, *supra* note 37, at 1633–34. It should be noted that a plaintiff does not necessarily need to prove the specific defect; rather, in some cases, a plaintiff could prove the defect by circumstantial evidence showing that the product malfunctioned. RESTATEMENT (THIRD) OF TORTS: PRODS. LIAB. § 3 (AM. LAW INST. 1998). For further discussion, see Geistfeld, *supra* note 37, at 1634–35 (suggesting that the product malfunction doctrine could apply to programming bugs in the context of autonomous vehicles); Owen, *supra* note 74, at 871–84.

79. RESTATEMENT (THIRD) OF TORTS: PROD.S. LIAB. § 2(b) (AM. LAW INST. 1998). This is referred to as a design defect.

80. Significant scholarship revolved around the question of whether the "risk-utility" or the "consumer expectations" approach prevailed. I will not attempt to resolve this dispute in this Article and will only mention that the risk-utility approach was favored by the Restatement (Third) of Torts: Products Liability, which explicitly replaced the consumer expectations approach manifested by the Restatement (Second) of Torts. *Id.* § 2 cmt. g. For elaboration, see Owen, *supra* note 75, at 360–67. *See generally* Richard A. Epstein, *Products Liability: The Search for the Middle Ground*, 56 N.C. L. REV. 643 (1978); Douglas A. Kysar, *The Expectations of Consumers*, 103 COLUM. L. REV. 1700 (2003); Aaron D. Twerski & James A. Henderson, Jr., *Manufacturers' Liability for Defective Product Designs: The Triumph of Risk-Utility*, 74 BROOK. L. REV. 1061, 1065 (2009).

reduced the risk imposed by the product using preventive measures that are reasonable in relation to the harm.[81] This is basically a cost-benefit analysis of alternative designs that imports a notion of negligence to the otherwise strict liability regime of products liability.[82] Note, however, that the risks and harms considered under this approach are only foreseeable risks and harms, and that the reasonableness of the design should not only be considered with respect to a specific harm done, but rather with respect to the product's safety at large.[83] A second approach to the design defect trigger is the "consumer expectations" approach, which asks whether the dangers imposed by a product exceed those reasonably expected by the potential consumers.[84]

The third trigger for products liability applies due to:

> inadequate instructions or warnings when the foreseeable risks of harm posed by the product could have been reduced or avoided by the provision of reasonable instructions or warnings by the seller or other distributor, or a predecessor in the commercial chain of distribution, and the omission of the instructions or warnings renders the product not reasonably safe.[85]

The main problem posed by AI-based products in the context of products liability—and mainly its second and third triggers—is the foreseeability problem discussed above. Both the design defect trigger and the inadequate warning or instruction trigger are focused on reducing foreseeable risks of harm.[86] While this may not be considered a problem with respect to some types of AI-based product, such as autonomous vehicles, the challenge is inevitable when it comes to fully AI-based products. Central to our idea of fully AI-based products is the expectation that such products perform, at least to a certain extent, in an unforeseeable manner. Had we wanted a product based on a pre-defined set of rules, we would not have equipped it with full AI capabilities. Most of the digital or connected products we use today rely on algorithms that do exactly that, operate on the basis of a predefined set of rules. The AI factor quantitatively changes the picture as it, by definition, leads to unexpected results.[87] Thus, if the unexpectedness is an inseparable part of the product and the consumer demands that it fulfills, it would be immensely difficult to base any form of liability on foreseeable or expected risks of harm. In fact, arguing that an AI-based product's design is defective due to its AI factor, or that users should have been warned or instructed with respect to the product's specific risky behavior (which is itself unforeseeable), means that many risks related to the unforeseeable aspects of AI should be foreseeable, thus leading to liability for AI-related harms in cases where the liable person cannot take adequate preventive measures or avoid the risk. The normative justifications for such a result are very difficult to establish,

81. *See* Owen, *supra* note 75, at 310.

82. See id.

83. *Id.* at 310–15.

84. *Id.* at 300–01; *see* RESTATEMENT (SECOND) OF TORTS § 402A cmts. g, i (AM. LAW INST. 1965).

85. RESTATEMENT (THIRD) OF TORTS: PRODS. LIAB. § 2(c) (AM. LAW INST. 1998).

86. *Id.* § 2 cmt. a ("Subsections (b) and (c) speak of products being defective only when risks are reasonably foreseeable.").

87. *See* Lemley & Casey, *supra* note 17, at 1324–26.

and in any event, this could not be supported by the common normative justifications to the design defect approach.[88] The contrary line of reasoning is more tolerable. Reaching the conclusion that, since AI-related risks are unforeseeable by nature and therefore cannot be covered by the design defect or duty of warning and instruction doctrines, there may be instances of harm lying outside the scope of products liability doctrine; but these cases may be compensated for by other forms of liability in torts as will be suggested below.[89]

An additional difficulty raised by AI-based robots in this context is the applicability of the concept of "product." As explained above, the AI factor of such products is based on software and algorithm.[90] Sometimes, such software and algorithms may be embedded in a product sold to the consumer and owned by it.[91] But this may not always be the case. In the age of the "access economy," many utilities that were once consumed as products purchased by the consumer are now delivered as services by a service provider.[92] The same may be true of various AI-based solutions, even household-related solutions that could be marketed and provided on an as-a-service model.[93] This shift from ownership to access may result in harms that do not enter the scope of the products liability doctrine.[94]

88. RESTATEMENT (THIRD) OF TORTS: PRODS. LIAB. § 2, cmt. a (Am. Law Inst. 1998) ("Subsections (b) and (c), which impose liability for products that are defectively designed or sold without adequate warnings or instructions and are thus not reasonably safe, achieve the same general objectives as does liability predicated on negligence. The emphasis is on creating incentives for manufacturers to achieve optimal levels of safety in designing and marketing products. Society does not benefit from products that are excessively safe—for example, automobiles designed with maximum speeds of twenty miles per hour—any more than it benefits from products that are too risky. Society benefits most when the right, or optimal, amount of product safety is achieved. From a fairness perspective, requiring individual users and consumers to bear appropriate responsibility for proper product use prevents careless users and consumers from being subsidized by more careful users and consumers, when the former are paid damages out of funds to which the latter are forced to contribute through higher product prices. . . . In general, the rationale for imposing strict liability on manufacturers for harm caused by manufacturing defects does not apply in the context of imposing liability for defective design and defects based on inadequate instruction or warning. . . . A reasonably designed product still carries with it elements of risk that must be protected against by the user or consumer since some risks cannot be designed out of the product at reasonable cost."). Another difficulty with such argument is that the opposite could equally apply. One could argue that if AI's risks are foreseeable, these should be expected by ordinary consumers who, by definition, should be aware of such risks and perhaps even assume them, leading to no liability on the manufacturer's part.

89. *See* discussion *infra* Sections III.B–C.

90. *See supra* notes 19–27 and accompanying text.

91. *See, e.g.*, Abraham & Rabin, *supra* note 38, at 141.

92. For various definitions of the access economy, see, for example, Steve Denning, *Three Strategies for Managing the Economy of Access*, FORBES (May 2, 2014, 10:22 AM), https://www.forbes.com/sites/stevedenning/2014/05/02/economic-game-change-from-ownership-to-access/#46b0195731c9; Giana M. Eckhardt & Fleura Bardhi, *The Sharing Economy Isn't About Sharing at All*, HARV. BUS. REV. (Jan. 28, 2015), https://hbr.org/2015/01/the-sharing-economy-isnt-about-sharing-at-all. For discussion on some legal implications of the access economy, see generally Orly Lobel, *The Law of the Platform*, 101 MINN. L. REV. 87 (2016).

93. For discussions on the effect of the internet of things on consumption and the shift from purchasing products to purchasing services, see Rebecca Crootof, *An Internet of Torts: Expanding Civil Liability tandards to Address Corporate Remote Interference*, 69 DUKE L.J. 583 (2019); Stacy-Ann Elvy, *Hybrid Transactions and the INTERNET of Things: Goods, Services, or Software?*, 74 WASH. & LEE L. REV. 77, 172 (2017).

94. This is obviously not to suggest that service providers do not bear any type of liability towards their consumers. They clearly do, in appropriate cases. The product liability doctrine, however, does not apply to harms caused by a service provider that is not selling or manufacturing a product involved in the provision of the services.

B. Abnormally Dangerous Activities

Following directly from the previous section, AI-based products could potentially fall under the category of tort law regulating abnormally dangerous activities. The Restatement (Second) of Torts provides that, "[o]ne who carries on an abnormally dangerous activity is subject to liability for harm to the person, land or chattels of another resulting from the activity, although he has exercised the utmost care to prevent the harm."[95] If we adopt the position arguing that the AI factor of such product is by definition unforeseeable (at least to a great extent), it may be argued that engaging in AI-related activity, whether by manufacturing, distribution, or operations, is abnormally dangerous, as is the case with various new and disruptive technologies whose risks are unknown. While this is an appealing proposition, it seems to be doctrinally difficult.

The Restatement provides six factors that courts should weigh to conclude whether an activity is abnormally dangerous: (i) the existence of a high degree of risk of some harm to a person or property; (ii) the likelihood that the harm that results from the activity will be great; (iii) the inability to eliminate the risk by the exercise of reasonable care; (iv) the extent to which the activity is not commonly used; (v) the inappropriateness of the activity to the place where it is carried out; and (vi) the extent to which its value to the community is outweighed by its dangerous attributes.[96] At first glance, it appears that most of the above factors could easily apply to AI-based products, at least under certain circumstances. But a more critical review could prove otherwise.

First, on a factual level, it may well be argued that factors (i) and (ii) do not apply. It seems very plausible that already today, and surely in the future, AI-based products will be much less likely to cause harms and that any such harms would not be of a greater magnitude than those harms caused by products operated and used by human beings.[97] Another factual argument could be made with respect to factor (iv). Clearly, at some point in the future, AI-based devices will be commonly used, perhaps even more than "un-intelligent" devices.[98]

Second, on a more normative level, factor (vi) seems inapplicable as well. It may well be argued that AI-based products are of great value to society, perhaps even exponentially higher than the risk they impose and the harms they will eventually inflict. As the Restatement puts it, "[t]his is true particularly when the community is largely devoted to the dangerous enterprise and its prosperity largely depends upon it."[99] In addition, due to the foreseeability problem that AI-based robots entail, it could be very difficult to prove the exact risks imposed by AI-based robots since, by definition, we simply cannot foresee many of them.[100]

95. RESTATEMENT (SECOND) OF TORTS § 519 (1) (AM. LAW INST. 1977).

96. *Id.* § 520.

97. For a similar argument in the context of autonomous vehicles, but not with respect to the abnormally dangerous activities doctrine, see Vladeck, *supra* note 6, at 146.

98. *2010s Decade in Review: The Rise of AI*, TECH TIMES (Jan. 3, 2020, 8:37 AM), https://www.techtimes.com/articles/246750/20200103/2010s-decade-in-review-the-rise-of-ai.htm.

99. RESTATEMENT (SECOND) OF TORTS § 520 cmt. k (AM. LAW INST. 1977).

100. This is much similar to the discussion on the design defect doctrine and the risk-utility test above.

Nevertheless, two factors may indeed lead to the applicability of this doctrine to AI-based robots. Factor (iii) seems to be applicable to almost all full AI-based robots and is to a certain extent the mirror-image of the policy considerations discussed with respect to factor (vi).[101] If we believe that AI is of significant benefit to society and that it entails risks that are unforeseeable by design, we must acknowledge that at least some risks associated with AI-based robots cannot be eliminated by the exercise of reasonable care.[102] But this factor alone cannot lead to the applicability of the doctrine to AI-based robots given that most of the other factors lead to the contrary conclusion.[103] It must be noted though that there could be circumstances in which factor (v) may apply, thus potentially subjecting an activity to this doctrine.[104] The concept of the locality of the activity referred to in factor (v) may apply when AI-based robots are deployed in environments that are unsuitable for such activity or where the potential harms are excessively high.[105] This could potentially apply to the use of fully AI-based medical equipment as the sole treatment of a patient without human supervision.[106]

C. Negligence

Negligence is the contemporary default liability rule in tort law.[107] It assumes that unintended harms or accidents should be compensated for by the injurer only if the injury is blameworthy or at fault.[108] The fault standard commonly accepted under this concept is that of a breach of a duty of care, namely the duty to act as a reasonable person.[109] There are generally four elements that should be considered in assessing whether an act of a person is negligent: the existence of a duty of care, a breach of such duty, harm caused to the victim, and a causal link between the breach and the harm caused.[110] Each of these elements has been extensively debated in legal scholarship, and I do not intend to offer

101. RESTATEMENT (SECOND) OF TORTS § 520 (AM. LAW INST. 1977).

102. *Id*. § 520 cmt. k ("What is referred to here is the unavoidable risk remaining in the activity, even though the actor has taken all reasonable precautions in advance and has exercised all reasonable care in his operation, so that he is not negligent. The utility of his conduct may be such that he is socially justified in proceeding with his activity, but the unavoidable risk of harm that is inherent in it requires that it be carried on at his peril, rather than at the expense of the innocent person who suffers harm as a result of it.").

103. *Id.* § 520 cmt. f ("In determining whether thedanger is abnormal, the factors listed in-Clauses (a) to (f) of this Section are all to be considered, and are all of importance. Any one of them is not necessarily sufficient of itself in a particular case, and ordinarily several of them will be required for strict liability.").

104. RESTATEMENT (SECOND) OF TORTS § 520 (AM. LAW INST. 1977).

105. *Id.* § 520 cmt. j.

106. Morris Panner, *How AI Supports and Accelerates Healthcare*, FORBES (Jan. 22, 2020, 8:10 AM), https://www.forbes.com/sites/forbestechcouncil/2020/01/22/how-ai-supports-and-accelerates-healthcare/#5312d07b1aa9.

107. RESTATEMENT (SECOND) OF TORTS § 282 (AM. LAW INST. 1965).

108. *See, e.g.*, Richard A. Posner, *A Theory of Negligence*, 1 J. LEGAL STUD. 29, 29 (1972).

109. RESTATEMENT (SECOND) OF TORTS § 283 (AM. LAW INST. 1965).

110. *See*, *e.g.*, David G. Owen, *The Five Elements of Negligence*, 35 HOFSTRA L. REV. 1671, 1672 (2007), also discussing other formulations of the basic elements of the doctrine such as duty, breach and proximately caused harm or a five element formulation of duty, breach, harm, cause, and proximate cause.

any new insights into such debates.[111] The purpose here is to generally portray the commonly accepted understanding of each element to show how it may or may not apply to risks caused by AI-based robots.

The duty element seems like a relatively low bar to pass, generally and with respect to AI-based robots in particular.[112] It has already been argued that courts generally acknowledge that an injurer has a general duty of care toward the victim unless there are strong policy considerations that determine otherwise.[113] It seems relatively safe to assume that, as a matter of policy, courts will generally determine that manufacturers, distributors, operators, and even users of AI-based robots have a general duty of care toward potential victims of the robots.[114]

The challenge becomes greater with the breach element. The most accepted formulation of the analysis for a breach of the duty of care is the Hand formula, determining that the injurer breaches the duty of care when (1) the costs of the prevention of harm were lower than the harm expectancy (the cost of harm multiplied by the probability of its occurrence), and (2) such preventive measures were not taken.[115] In the context of AI-based robots, each factor in this formula becomes very difficult for the potential tortfeasors to know *ex ante* and no less difficult for courts to determine *ex post* what the tortfeasors should have known at the time. If the behavior of AI-based robots is unpredictable or unexplainable to humans, how can the probability that they will eventually inflict harm on others be assessed? Moreover, the harm itself is not only unforeseeable or difficult to assess, but also the types of harm themselves are sometimes unpredictable and new in nature.[116] On the other side of the equation, if all of the above is highly unforeseeable and unexplainable, are there even preventive measures that could be taken by designers, operators, or end-users of AI-based robots? And in the case that measures are taken, how can we possibly assess them without resorting to binary discussions of pursuing or not pursuing the activity at large?[117]

111. *Id*. at 1672–73.

112. RESTATEMENT (THIRD) OF TORTS: PHYS. & EMOT. HARM § 7 (AM. LAW INST. 2010).

113. For a discussion on main arguments for duty skepticism and replies to such arguments, see John C. P. Goldberg & Benjamin C. Zipursky, *The Restatement (Third) and the Place of Duty in Negligence Law*, 54 VAND. L. REV. 657, 692–97 (2001).

114. The duty question could become more complex when considering whether manufacturers have a duty of care towards third parties with whom users are interacting through the use of AI-based robots they operate, or in other cases where there is a chain of stakeholders that could break the duty chain. I believe, however, that this complexity better fits the general discussion on causation and foreseeability rather than the mere existence of a duty of care. For a discussion of cases in which no-duty is a viable option, see *id.* at 665–74.

115. This is an informal representation of the algebraic Hand formula of B<PL, as determined in United States v. Carroll Towing Co., 159 F.2d 169, 173 (2d Cir. 1947). For a more formal economic formula of due care, see William M. Landes & Richard A. Posner, *The Positive Economic Theory of Tort Law*, 15 GA. L. REV. 851, 852 (1981). *See also* RESTATEMENT (SECOND) OF TORTS §§ 283, 291–93 (AM. LAW INST. 1965).

116. John Loeffler, *AIs Continue to Act in Unpredictable Ways, Should we Panic?*, INTERESTING ENGINEERING (Jan. 4, 2019), https://interestingengineering.com/ais-continue-to-act-in-unpredictable-ways-should-we-panic.

117. It is apparent that this discussion revolves mainly around the idea of the inherently unforeseeable nature of AI-based robots. While it is debatable whether foreseeability is part of the breach element of negligence, *i.e.*, whether the harm and probability of harm really refer only to foreseeable harms and not to all harms, I believe that these difficulties apply to the breach element in the context of AI-based robots without formally reading

Considering the causation element of negligence, eminent issues arise as well. First, we must again pass the personhood and agency problem of robots.[118] For human tortfeasors to be liable for negligence, their acts should be legally linked to the harms caused.[119] AI-based robots sometimes act in a manner that cannot be comparably described to the acts of a human. But even if we extend the causation concept using the principal-agent relationship or by assuming that the mere choice to design, operate, or use an AI-based robot may be a legal cause for harms inflicted by such robots, we are still faced with the question of foreseeability of harms and types of harms that may be caused.[120] The lack of foreseeability and perhaps even the inexplicability of AI-robot behavior could, in many cases, break the causation link between the human who interacts with the robot and the victim of the robot's conduct.

IV. TORT LAW OBJECTIVES AND ALTERNATIVE AI TORT REGIMES

After establishing that current tort doctrines struggle with the special characteristics of AI-based robots, we can move away from specific doctrines to more general ideas of tort liability and assess which liability regime should apply to risks posed by AI-based robots. For this, we must first briefly refresh our memories of the basic principles of tort law.

The objectives of tort law—the law of accidents—are generally characterized, at least from an economics viewpoint, as promoting safety rules that will lead to socially optimal costs of losses caused by accidents and safety measures taken to prevent such losses.[121] The promotion of such safety rules is often referred to as *deterrence* and *allocation of resources*.[122] Economic literature on torts typically focuses on choosing the best liability rules or regimes to achieve these goals.[123] The main alternatives are strict liability and negligence, to which certain complexities such as contributory or comparative negligence rules are added.[124] These are usually considered in cases of unilateral accidents and bilateral accidents.[125]

foreseeability into the definition of breach. For the assertion that foreseeability is adopted by courts as part of the breach element, see Zipursky, *supra* note 61, at 1255–57.

118. *See* discussion *supra* Section II.B.

119. RESTATEMENT (THIRD) OF TORTS: PHYS & EMOT. HARM §§ 26, 29 (AM. LAW INST. 2010).

120. For a discussion on why foreseeability is necessarily embedded in the idea of causation as an element of negligence, see Zipursky, *supra* note 61, 1266–71.

121. Jules L. Coleman, *The Structure of Tort Law,* YALE L.J. 1233 (1987–1988) (reviewing books).

122. STEVEN SHAVELL, ECONOMIC ANALYSIS OF ACCIDENT LAW 297–98 (2009); Gary T. Schwartz, *Mixed Theories of Tort Law: Affirming Both Deterrence and Corrective Justice*, 75 TEX. L. REV. 1801, 1803–06 (1997). There are, of course, other objectives aimed at achieving the optimal equilibrium, such as administrative costs and distributive considerations. SHAVELL, *supra,* at 262–65. I will discuss the effects of administrative-cost considerations below.

123. *See generally* Landes & Posner, *supra* note 115.

124. A third alternative to no-liability is often considered to demonstrate an inefficient condition. *See* SHAVELL, supra note 122, at 8.

125. Unilateral accidents are accidents that only the injurer could avoid, whereas bilateral accidents could be avoided by measures of care taken by both injurer and victim. *See*, *e.g.*, *id.* at 6–7, 9–10.

A. Strict Liability

Under a strict liability model, actors will be liable for the risks associated with their acts regardless of whether they were at fault.[126] As Gregory Keating put it, "fault liability makes *wrongful* agency the fundamental basis of responsibility for harm accidentally done; strict liability makes *agency* itself the fundamental basis of responsibility."[127] Historically, strict liability was the general tort rule holding that people should act at their peril.[128] But as fault-based liability under the idea of negligence became the default liability rule in tort, commentators attempted to theorize the concept of strict liability and set the rules for its application to certain circumstances.[129]

The common understanding of the applicability of strict liability, at least under an economics-centered approach to tort law, is that strict liability should apply when the injurers are better situated to determine the costs of risk associated with their actions.[130] As Guido Calabresi and Jon Hirschhoff phrased it, liability under strict liability should lie with the party who "is in the best position to make the cost-benefit analysis between accident costs and accident avoidance costs and to act on that decision once it is made."[131] This was the basis, for example, for the early products liability doctrine.[132] The presumption was that the manufacturer of a product is better situated to assess the risk associated with the product, as well as to take preventive measures, if necessary, to mitigate such risk, than are the consumers.[133] Cases in which consumers use products against the manufacturer's instructions or in ways unexpected by the manufacturer would be exceptions to the strict liability rule.[134]

One of the first difficulties with applying strict liability theory to AI-based robots revolves around the personhood and agency problems discussed above. Although strict liability does not focus on fault, and in this sense may be able to capture unexpected behavior of AI-based robots from the duty standpoint (in contrast to negligence), it still requires foreseeability with respect to the harms caused on the causation side and the manifestation of agency or volition by the

126. Richard A. Epstein, *A Theory of Strict Liability*, 2 J. LEGAL STUD. 151, 152 (1973).

127. Gregory C. Keating, *The Theory of Enterprise Liability and Common Law Strict Liability*, 54 VAND. L. REV. 1285, 1286 (2001).

128. *See, e.g.*, Epstein, *supra* note 126, at 152–53; Keating, *supra* note 127, at 1287; Posner, *supra* note 108, at 29.

129. I will by no means attempt to settle the divide between scholars at large—particularly law and economics theorists and those focused on rights-based approaches—on whether strict liability is indeed an exception to the general rule of negligence or is a general competing theory of tort. My attempt here is to briefly describe the common understanding of strict liability as a positive liability rule.

130. SHAVELL, *supra* note 122, at 8.

131. Guido Calabresi & Jon T. Hirschoff, *Toward a Test for Strict Liability in Torts*, 81 YALE L.J. 1055, 1060 (1972). Another approach to strict liability holds that its application should depend on concepts of fairness, based on the idea that enterprise liability should apply when actors benefits from the risk they create. *See generally* Keating, *supra* note 127. Other rights-based approaches attempt to fit the idea of strict liability within the fundamental understanding of tort law as a correlative framework of rights and duties. *See*, *e.g.*, ERNEST J. WEINRIB, CORRECTIVE JUSTICE 1–8 (2012).

132. *See, e.g.*, Calabresi & Hirschoff, *supra* note 131, at 1056.

133. *Id.* at 1062.

134. *Id.* at 1061–64.

actor.[135] This point returns us to the general problems of the lack of agency and personhood of robots and the difficulty in linking human stakeholders with the risk-inflicting behavior of robots, absent liability extension doctrines such as principal-agent relationships.[136]

But more substantively, even if we take the extra step and create a link between human stakeholders and certain actions of AI-based robots, which human would be better situated to make the cost-benefit analysis of risks and their avoidance? In contrast to the simpler case of products liability, where manufacturers could be assumed to better understand the risks involved in using their products and perhaps even prevent them, this may simply not be the case for AI-based robots. The general problem of the lack of foreseeability of robot behavior is crucial to make the normative decision regarding the allocation of risk under strict liability. In some cases, one stakeholder may have more information than other stakeholders with respect to potential risks.[137] For example, operators of AI-based robots in the medical device field may generally have more knowledge regarding the risks involved in the procedure for which the robot is used than the user-patient. The opposite may be true when users are operating a general-purpose AI-based personal assistant robot to monitor their children when leaving them home alone. But since the main feature of AI-based robots is that they may act in a manner unforeseeable or unexplainable by humans, in many cases none of the stakeholders will be better situated to assess the risks involved in their operation and the problem of imperfect information will apply equally to all stakeholders.[138] In such cases, the guiding applicability principles for strict liability cannot apply.[139]

135. *See, e.g.*, Smith, *supra* note 38, at 37.

136. *See supra* Section II.B.

137. It is well accepted that from an economic analysis perspective, strict liability is efficient in bilateral accidents only if a contributory negligence rule applies, meaning that if the victim does not demonstrate due care, she cannot recover from the injurer the costs of the accident. *See, e.g.*, SHAVELL, *supra* note 122, at 15. In some cases involving of AI-based robots, given the lack of immediate agency or connection to one specific person, accidents may be considered tri-partite (or more), or at least it would be very difficult to determine who the applicable cost-avoiders are.

138. Steven Shavell explains that choosing between negligence and strict liability when both parties have perfect information is irrelevant. Steven Shavell, *Strict Liability Versus Negligence*, 9 J. LEGAL STUD. 1, 6 (1980). He further noted that when victims' knowledge is imperfect and injurers'' knowledge is perfect, a strict liability model will be favorable. *See id*.; *see also* SHAVELL, *supra* note 122, at 53–54. I believe that, conversely, if both parties suffer from equal lack of information on the risks, a strict liability model will be inefficient, or no more efficient than alternative liability regimes. One of Alan Schwartz's arguments against strict liability in the case of products liability is that the problem of the consumers' imperfect knowledge of risk is remediated by a liability rule that creates an imperfect knowledge problem for sellers due to their inability to assess the non-pecuniary losses awarded to consumers. *See* Alan Schwartz, *The Case Against Strict Liability*, 60 FORDHAM L. REV. 819, 834 (1992). The same logic applies when the imperfect information problem is equally inherent to both parties to start.

139. Under fairness-based approaches, it may be argued that manufacturers and distributors of AI-based robots should nevertheless be strictly liable for any risk associated with their activity, since they profit from this activity. Yet, this is not necessarily accurate. Users may make more gain operating AI-based robots than the manufacturers. Moreover, in user-to-user engagements, the entire equation changes in this regard. For the applicability of fairness considerations in this context, see *infra* Section V.D.

It could be argued that strict liability will require designers to obtain the necessary information to make the optimal cost-benefit analyses of their activity.[140] But in the context of AI-based robots, this may not be technologically possible. In such cases where unpredictability is embedded in the technology, there is no advantage to imposing strict liability for the purpose of creating foreseeability. On the contrary, in these circumstances, such a liability regime may lead to suboptimal activity levels of designers due to the lack of information.

B. Negligence as a Liability Regime

I previously explained the shortcomings of negligence as the main tort law doctrine. But negligence is also one of the common alternative liability regimes in tort.[141] Theoretically, we could conclude that negligence would be the optimal liability regime in the context of AI-based robots although the current negligence doctrine is not optimal. Thus, we should briefly discuss the optimality of negligence as a liability regime in our context.

The common perception in economic literature is that in cases of bilateral accidents, negligence, in all forms, would generally be socially optimal.[142] This is because injurers will be induced to take due care to refrain from bearing the costs of accidents, and victims will take due care to avoid risks to lower their costs in cases of accidents for which injurers are not legally responsible.[143] But the crux of the economic argument for the optimality of negligence is that courts must set an optimal level of due care.[144] As explained above, in the contexts of the negligence tort doctrine and AI-based robots, setting an optimal level of due care seems to be a difficult task—or impossible in some circumstances.[145] In many cases, the foreseeability problem of AI will make it extremely difficult to determine what the optimal precautions that persons linked to AI-based robots should take are and what safety measures potential victims engaging with such robots should take. When courts are unable to determine optimal due care, negligence becomes a sub-optimal liability regime for two main reasons. First, injurers and victims are not optimally deterred from neglecting precautions, and second, the already existing administrative costs of courts adjudicating due care increase.[146]

C. No-Fault Mandatory Insurance

A final tort-like mechanism that may be used to compensate for injuries caused by AI-based robots is a no-fault mandatory insurance scheme. The automotive field primarily adopted no-fault mandatory insurance to compensate for

140. *See generally* Steven Shavell, *Liability and the Incentive to Obtain Information About Risk*, 21 J. LEGAL STUD. 259 (1992) (arguing that strict liability will always cause injurers to obtain information optimally).

141. *See, e.g.*, SHAVELL, *supra* note 122, at 8.

142. *See*, *e.g.*, *id.* at 9–15.

143. *Id.*

144. *See id.* at 16.

145. *See supra* Section III.C.

146. *See* SHAVELL, *supra* note 122, at 15–18, 264.

bodily injuries resulting from accidents, regardless of fault.[147] It is based on a mandatory duty to purchase first-party insurance and a restriction on the right to sue the injurer.[148] It is mandatory because it is intended to completely replace the tort system for such injuries and refrain from resorting to questions of negligence or fault.[149] No-fault insurance schemes have various economic advantages such as reducing administrative costs and producing more equitable outcomes, as victims are compensated regardless of their ability to prove negligence or fault.[150] Following the shortcomings of existing tort law doctrines with respect to AI-based robots as discussed above, this seems like a potentially effective solution.

In fact, the European Parliament advised the European Union Commission to consider and adopt a mandatory insurance scheme with respect to robotics and AI.[151] The recommendation consisted of five general principles: (i) establishing a mandatory insurance scheme for specific categories of robots and requiring the producers of the robots to purchase such insurance; (ii) establishing a compensation fund granting monetary compensation to victims of robot behavior; (iii) limiting the liability of robot producers if they contribute to such compensation fund and purchase insurance; (iv) selecting between general compensation funds or sector-specific funds; and (v) establishing a robot registry that will establish a valid link between a robot and the fund with which it is associated.[152]

A no-fault insurance scheme takes much of the sting out of the shortcomings of the abovementioned doctrines in the context of AI-based robots. Since this is a no-fault regime, questions of foreseeability, personhood, and liability, which were the core difficulties in the AI context, are taken out of the equation, clearing the road for a potentially consistent liability model for AI-based robots. But the idea of a mandatory insurance scheme protecting against risks associated with AI-based robots is not free from concerns, some of which undermine the entire concept.

First, we must consider the main criticism of no-fault regimes at large. Legal and economic literature suggests that, while no-fault regimes may be more efficient due to their savings in administrative costs and judicial errors, they may

147. Several United States jurisdictions, several provinces of Canada and Australia, New Zealand, and Israel have adopted such schemes. *See* JAMES M. ANDERSON ET AL., THE U.S. EXPERIENCE WITH NO-FAULT AUTOMOBILE INSURANCE: A RETROSPECTIVE 11–12 (2010); Craig Brown, *Deterrence in Tort and No-Fault: The New Zealand Experience*, 73 CALIF. L. REV. 976, 976 (1985); J. David Cummins et al., *The Incentive Effects of No-Fault Automobile Insurance*, 44 J.L. & ECON. 427, 427 (2001); Yu-Ping Liao & Michelle J. White, *No-Fault for Motor Vehicles: An Economic Analysis*, 4 AM. L. & ECON. Rev. 258, 258 (2002).

148. Cummins et al., *supra* note 147, at 427. Note that while the duty is usually with the potential injured party (in this case the driver), the costs of covering insurance premiums could be otherwise distributed to, for example, manufacturers or other operators of a service to which a mandatory insurance scheme applies.

149. ANDERSON ET AL., *supra* note 147, at 14.

150. Liao & White, *supra* note 147, at 259.

151. European Parliament Resolution of 16 February 2017 with Recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)), EUR. PARL. DOC. P8_TA (2017) 0051 (2017).

152. *Id.* at 15–16.

increase the number of accidents due to lack of deterrence.[153] Several commentators have attempted to prove or disprove this theoretical assumption.[154] In the context of AI-based robots, it is questionable whether this is even a factor. As robots are not currently given legal personhood and that their behavior is, at least to some extent, unforeseeable to their designers, operators, and users, it is questionable whether the concept of deterrence is even relevant. But if we do assume, or at least aspire, that our liability concepts will have some ex ante effects on behavior, whether with respect to human stakeholders or the robots themselves, we must take into consideration the effect no-fault regimes may have on such deterrence.

Second, and perhaps more importantly, since a mandatory insurance scheme is intended to fully supplant the general tort system, it requires a hermetic and wide adoption by all stakeholders.[155] This was politically difficult to achieve in a non-AI-robots world. It is no surprise that such models were adopted in the automotive context which is, by nature, relatively local and physically-bound.[156] But this becomes even more difficult in the context of robotics. The physical-digital nature of many robots makes it almost impossible to have a single rule that will apply to all stakeholders. The manufacturer of the robot could be American, the operator British, and the end-user Japanese. For a perfect mandatory insurance scheme to be feasible, all relevant jurisdictions must adopt it, a task that seems politically impossible. This is not to say that a mandatory insurance model could never work for risks associated with AI-based robots. We can definitely consider such a model in the context of autonomous vehicles, which are quite similar to regular cars in this context.[157] It could also be feasible in the context of medical robots, at least in jurisdictions that have mandatory health insurance or collective health benefits.[158] But my point here is that the lack of physical borders in many circumstances related to AI-based robots and the political impracticability of adopting global or cross-jurisdictional mechanisms make

153. *See*, *e.g.*, ANDERSON ET AL., *supra* note 147, at 83–86; Brown, *supra* note 147, at 976–79; Cummins et al., *supra* note 147, at 427.

154. Cummins et al., *supra* note 147, at 427; Rose Anne Devlin, *Some Welfare Implications of No-Fault Automobile Insurance*, 10 INT'L REV. L. & ECON. 193, 202 (1990); Elisabeth M. Landes, *Insurance, Liability, and Accidents: A Theoretical and Empirical Investigation of the Effect of No-Fault Accidents*, 25 J.L. & ECON. 49, 49 (1982); R. Ian McEwin, *No-Fault and Road Accidents: Some Australasian Evidence*, 9 INT'L REV. L. & ECON. 13, 14 (1989).

155. *See generally* Landes, *supra* note 154.

156. For New Zealand and Israel, which have no land borders with any other jurisdictions, such a model makes perfect sense because it could be inherently enforced with respect to all stakeholders. In regions where vehicles could cross physical borders and jurisdictions, such as the United States, this solution becomes more complex.

157. This is actually suggested by Kenneth Abraham and Robert Rabin as a more efficient solution when all car accidents are caused by autonomous vehicles. *See* Kenneth S. Abraham & Robert L. Rabin, *Automated Vehicles and Manufacturer Responsibility for Accidents: A New Legal Regime for a New Era*, 105 VA. L. REV. 127, 145–47 (2019).

158. For a suggestion as to how to apply the no-fault regime in the context of medical malpractice, see Frank A. Sloan et al., *The Road from Medical Injury to Claims Resolution: How No-Fault and Tort Differ*, 60 LAW & CONTEMP. PROBS. 35 (1997).

the no-fault model irrelevant as a *general* solution to the shortcomings of tort law doctrines in this context.

Third, even if we believe that a no-fault model could apply to all or some risks related to AI-based robots, the questions of cost allocation and premium estimation become much more difficult to address in the context of AI-based robots. As explained above, the multi-stakeholder feature of AI-based robots raises the question of who should pay for such insurance costs. In the current case of automobiles, drivers (and passengers) are largely both the tortfeasors and victims of car accidents. It is therefore relatively easy to determine that drivers should have a duty to purchase mandatory insurance policies covering such risks, assuming that there is a balanced cross-subsidy between drivers. But when AI-robots are governed by their designers, operators, and users, the question of who pays insurance premiums becomes more complex.[159] In fact, it is expected that most accidents involving autonomous vehicles will result from human behavior (not necessarily of the driver, rather often of pedestrians).[160] Mandatory insurance for drivers or manufacturers will impose liability and deter the wrong tortfeasors. If that is not enough, the augmented harms that characterize AI-based robots, as well as the un-foreseeability problem, are destined to make the task of determining insurance premiums almost impossible.

V. A NEW LIABILITY MODEL FOR AI: PRESUMED NEGLIGENCE WITH SAFE HARBORS

After reviewing the shortcomings of current tort law doctrines in the context of potential liability for AI-based robots and discussing the adequate liability regime, we can now turn to ask what could be done and what rules could be adopted to adequately adapt tort law to capture harms that are caused by AI-based robots and that should be compensated for, under a proper liability rule.[161]

159. The European Parliament suggested that manufacturers or owners of robots bear the cost of insuring against potential injuries. *See* European Parliament Resolution of 16 February 2017 with Recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)), EUR. PARL. DOC. P8_TA (2017) 0051 15 (2017). But this suggestion seems arbitrary. It may be more efficient to divide the costs of insurance between owners and users or have only users or only manufacturers pay for insurance, all based on various factors to be considered.

160. Don Reisinger, *Humans—Not Technology—Are the Leading Cause of Self-Driving Car Accidents in California*, FORTUNE (Aug. 29, 2018, 10:17 AM), https://fortune.com/2018/08/29/self-driving-car-accidents/.

161. Vladeck asserted that any type of harm caused by AI should be compensated for; otherwise users and consumers will be left without proper compensation for their injuries. *See* Vladeck, *supra* note 6, at 128. I disagree with this assertion. The question of whether specific harm or injuries should be compensated for could be answered only on the basis of our normative understanding of the grounds for liability in tort. Regardless of our guiding theory, such assertion cannot always be true. Under the economic analysis approach, harms should be compensated where it is efficient to do so under a cost-benefit analysis—for example in order to internalize negative externalities or in order to deter wrongdoers from doing wrong in the first place. This could happen when the person inflicting harm is in a better position to prevent it from happening and therefore should be deterred and internalize her externalities. *See*, *e.g.*, SHAVELL, *supra* note 122, at 297–98; Guido Calabresi & A. Douglas Melamed, *Property Rules, Liability Rules, and Inalienability: One View of the Cathedral*, 85 HARV. L. REV. 1089 (1972); Lemley & Casey, *supra* note 17, at 1353–56. Under a corrective justice approach, an injury must be compensated for only if the injurer has a correlative duty to refrain from inflicting the harm to begin

As several commentators have recently suggested, acknowledging some sort of culpability or agency of machines and robots could potentially resolve the shortcomings presented above.[162] Mark Lemley and Bryan Casey wrote a fascinating review of the challenges of imposing and enforcing remedies on robots (and perhaps their designers and operators) and suggested principles for aligning remedies for robots with our general normative justifications for remedies at large.[163] Other commentators have suggested imposing regulatory rules with respect to the coding and design of robots and autonomous products.[164] Matthew Scherer has even suggested establishing a regulatory authority dedicated to regulating and governing the development of AI.[165]

In this Part, I suggest taking a different path that may resolve the shortcomings of existing tort law doctrines, at least until the question of agency and personhood of robots can be revisited, presumably a decade or more from now. Instead of resorting to conceptually new models of remedies and liability, I suggest enhancing an existing liability rule, namely negligence, with supplementary rules that will set a predetermined acceptable level of care applicable to designers and operators of AI-based robots (regardless of whether AI is embedded in the product sold to the consumer or AI capabilities are delivered as a service).[166] Such level of care or quasi-safe-harbors, if unmet, will trigger liability by designers, operators, or end-users of AI-based robots in a manner that is not challenged by the basic unique problems related to AI, such as foreseeability and agency, effectively forming a presumption of negligence.[167] In contrast, if the rules are met,

with, under a right-based understanding of rights and duties. *See* WEINRIB, *supra* note 131, at 2–10; Omri Rachum-Twaig & Ohad Somech, *The Right-Based View of the Cathedral: Liability Rules and Corrective Justice*, 2016 PEPP. L. REV. 74, 80–81 (2016); Lemley & Casey, *supra* note 17, at 1353–56. Under both theories, some injuries will eventually be borne by the victim.

162. Lemley & Casey, *supra* note 17, at 1378–1386; Vladeck, *supra* note 6, at 150. Bryan Casey added a utopian vision according to which morality could be coded to robots' behavior, thus resolving the difficulties with imposing liability on robots. *See* Casey, *supra* note 64, at 1365. While this may actually happen in the not-so-far future, since this Article focuses on cases of AI-based robots, it seems extremely difficult if not impossible to ensure that the morality coding of such machines will curtail their independent behavioral capabilities.

163. Lemley & Casey, *supra* note 17.

164. This was largely suggested in the context of autonomous vehicles. *See, e.g.*, Abraham & Rabin, *supra* note 157, at 136; Geistfeld, *supra* note 37, at 1693; Bryant Walker Smith, *Automated Driving and Product Liability*, 2017 MICH. ST. L. REV. 1, 74 (2017).

165. Matthew U. Scherer, *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies*, 29 HARV. J.L. & TECH. 353, 393–97 (2016).

166. As explained above, one of the shortcomings of the product liability doctrine is that it does not apply to service providers. Given that many AI capabilities are already, and will increasingly be, deployed on an as-a-service model, any adequate solution to the problem must include service providers as potential culpable tortfeasors. For a similar suggestion in the context of the Internet of Things, see Crootof, *supra* note 93, at 40 (suggesting the companies distributing IoT devices should be considered IoT fiduciaries and thus have a duty of loyalty to their consumers). *See also* Jack M. Balkin, *Information Fiduciaries and the First Amendment*, 49 U.C. DAVIS L. REV. 1183, 1186 (2016) (suggesting the concept of "information fiduciaries" to explain the special relationship between companies that collect significant amounts of data from their consumers and users and the users themselves, a relationship that may impose quasi duty of loyalty on such companies).

167. For an analysis of how the products liability doctrine is, to a certain extent, a presumed negligence standard, see David P. Griffith, *Products Liability—Negligence Presumed: An Evolution* , 67 TEX. L. REV. 851, 853 (1989).

specific tort doctrines such as products liability will not apply, and the basic negligence rule will apply, but plaintiffs will have to prove actual negligence, forming a quasi-safe-harbor. At least in some circumstances, however, given the information generated by meeting such supplementary rules, the task of proving negligence may become more feasible in comparison to the current doctrine.

The full theoretical justification for such model and its effect on traditional analyses of the alternative tort liability regimes will be discussed elsewhere. In a nutshell, however, the purpose of the suggested rules is to put the relevant stakeholders in a condition allowing them to be aware of the risks posed by AI-robots and the ways to protect against them, effectively making them meet a socially acceptable level of care. These safe harbors will apply to those stakeholders that are better situated to implement them (much like a strict liability rule), but liability itself will apply only if the supplementary rules are unmet or if circumstances arise, after meeting such rules, that justify liability under existing tort doctrines.[168]

I offer the following framework as an alternative in those circumstances in which AI-based robots cause harms characterized by the agency and foreseeability problems in a manner that disrupts current tort doctrines and liability regimes. This is not to say that no such harms could be resolved by current doctrine. As explained above, there may be cases in which AI-based robots behave in a manner that does not raise the agency or foreseeability problems (or both). For such cases, current doctrine should prevail.

A. Monitoring Duty

As explained above, while designers and operators of AI-based robots are not well situated to assess their associated risks due to problems of un-foreseeability and inexplicability, they may be better situated to employ and embed monitoring technologies that could alert or signal to them or any other stakeholder when something goes wrong. The challenge here is to implement monitoring tools that will not undermine the basic purpose of the AI-based robot. It would, of course, be easier, and perhaps safer, to implement tools that simply disable certain functionalities or conduct of the robot to the extent that these are unexpected. But this goes against the basic idea of AI robotics. Superior monitoring could be achieved by implementing anomaly-based monitoring systems programmed to give warning when a robot behaves in an unexpected manner.[169] Other monitoring technologies could be based on AI themselves, studying the

168. This is an economic formulation of a justification as to why we would like to impose such duties on the relevant stakeholders. A different formulation could draw from Balkin's information fiduciaries concept and Crootof's extension of the concept to the idea of IoT fiduciaries. *See* Balkin, *supra* note 166, at 1186; Crootof, *supra* note 93, at 39.

169. Anomaly-based monitoring is commonly used in the field of cyber security to defend against unpredictable or "zero day" attacks. *See* P. Garcia-Teodoro et al., *Anomaly-Based Network Intrusion Detection: Techniques, Systems and Challenges*, 28 COMPUTERS & SECURITY 18, 19 (2009). The same idea could be adopted for AI-based robot monitoring.

tendencies of a specific robot and predicting whether it will behave in an unexpected harm-inflicting manner.

Once such monitoring is implemented, a duty to inform potential victims of the robot immediately follows.[170] The merit in implementing such monitoring and notification tools lies first and foremost in its potential to prevent harm. But it may also serve as a basis for determining more proper liability rules for categories of cases. For example, it may be that after employing monitoring abilities, the designer or operator of an AI-based robot will be better situated than the potential victim to assess risks and take preventive measures, supporting the imposition of strict liability. And even if there is still no basis for strict liability, such monitoring abilities may lead to circumstances in which a negligence analysis would be viable with respect to the designer's or operator's actions, and perhaps also for the assessment of the victim's contributory negligence. In addition, a monitoring duty may lead, in some circumstances, to a sounder application of the principal-agent relationship between the monitoring stakeholder and the robot, establishing the control element of vicarious liability.[171]

It must be noted that such monitoring duties may rightfully raise privacy concerns. I obviously do not recommend solving the problems of AI liability by causing additional privacy risks. Such monitoring duties will have to be accompanied by privacy-by-design duties allowing for behavior monitoring of the robot without exposing any personal data of individuals with whom the robot interacts.[172]

B. Mandatory Emergency Brakes

Another aspect with respect to which designers of AI-robots are better situated is the potential ability to include emergency brakes, shut-down capabilities, or features that make a robot unintelligent at the press of a button. This may not always be possible, and under some dystopic predictions robots may be able to circumvent such features, but as a general idea, we may require AI-based robots' designers to include such features at the design stage. The merits of this supplementary duty are mainly preventive. If something is about to go wrong, the operator or user of an AI-based robot could simply shut it down. When monitoring duties are imposed, the manufacturer or operator may even be required to remotely shut down the robots themselves.

Including the shut-down feature in the design does not mean that the designer or operator is exempt from liability, nor that this person will eventually

170. This is similar to a post-sale duty to warn under the products liability doctrine. *See* RESTATEMENT (THIRD) OF TORTS: PRODS. LIAB. § 10 (AM. LAW INST. 1998).

171. *See* RESTATEMENT (THIRD) OF AGENCY §§ 7.03–7.07 (AM. LAW INST. 1998); *see also supra* notes 52–59 and accompanying text.

172. This could be done by anonymization of monitored data or monitoring at network levels that do not expose the human-readable data itself. For elaboration on privacy-by-design, see ANN CAVOUKIAN, PRIVACY BY DESIGN IN LAW, POLICY AND PRACTICE (2011). For an example of regulatory rules requiring privacy-by-design, see Commission Regulation 2016/679, 2016 O.J. (119) 1 (EU) (regarding the protection of natural persons with regard to the processing of personal data and on the free movement of such data).

have liability if harms occur. In some circumstances, it would be the user or operator who is be better situated to cease the operation of the robot. In other circumstances, it may be the monitoring designer or distributor. In certain instances, it would not at all be reasonable or efficient to cease the operation of the robot, in which case no liability would apply. On the contrary, a careless shut-down of a robot, especially if done by a remote operator or distributor, may be negligent on its own account. It could be established, however, that not including such shut-down feature at all would be considered a design defect under the products liability doctrine.[173]

C. Ongoing Support and Patching Duties

A duty parallel to a general monitoring duty could be an ongoing support and patching duty. This, to some extent, resembles the post-sale duties of warning and instruction and the duty to recall defective products under the current products liability doctrine.[174] The difference here, however, is that general inferences from one revealed "defective" behavior of an AI-based robot to another may not be possible. This is because, as explained above, the learning process of AI differs according to the data and interactions the robot has, which may vary post-sale and distribution. But equipped with abnormality-based monitoring technologies, as may be required under a monitoring duty, designers and distributors of AI-based robots may be able to make statistical or case-specific inferences from monitoring results that could be translated to a duty to recall robots and patch them, to the extent possible. Here, there is no doubt that if anyone were in a position to understand the risk and prevention methods, it would be the designers of the robots, and, if enough information is obtained by them, a breach of a duty to support and patch robots by means of recalls may lead to liability for harms caused at a later stage.

D. Who Should Be Liable? Identifying the Stakeholders

But which stakeholder should be presumed negligent? As with any case of potential liability in tort, one important task is identifying the relevant stakeholders that could be considered liable. While the plaintiff will usually identify one or more defendants, it would be socially optimal to have all potential tortfeasors considered, due to the basic tort rules regarding multiple tortfeasors and whether liability is joint or several.[175]

173. *See* RESTATEMENT (THIRD) OF TORTS: PRODS. LIAB. § 2(b) (AM. LAW INST. 1998); *see also supra* notes 73–94 and accompanying text.

174. *See* RESTATEMENT (THIRD) OF TORTS: PRODS. LIAB. § 11 (AM. LAW INST. 1998).

175. *See, e.g.*, RESTATEMENT (SECOND) OF TORTS § 875 (AM. LAW INST. 1979) (stating that tortfeasors causing indivisible harm are jointly liable for such harm); *id.* § 881 (stating that tortfeasors causing divisible harms are only severally liable for such harms); RESTATEMENT (THIRD) OF TORTS: APPORTIONMENT OF LIAB. §§ 10–21 (AM. LAW INST. 2000).

1. Manufacturers (Designers)

Manufacturers, or designers in our case, are usually held accountable for some problems caused by their products.[176] This is because they usually have better knowledge and information on the characteristics and features of their products, as well as the ability to control both their safe manufacturing and design. But this may not be the case with AI-based robots. Surely, designers have better knowledge than perhaps any other stakeholder on how to create a robot, using machine-learning technologies. But at least with respect to the distinct features of AI, and mainly the unforeseeable behavior of the robot throughout its life, designers may have the same knowledge and information other stakeholders have.

As we will see, this distinct feature will have us question whether and to what extent we are comfortable holding designers accountable for harms inflicted by robots that they designed. At the very least it shakes the ground of various tort liability doctrines, specifically those based on the idea of strict liability. Moreover, in some instances, as shall be discussed below, AI-based robots may not at all be considered products, at least not in the common seller-consumer sense, and therefore the role of designers may not be as significant as that of other stakeholders involved in the interaction.

2. Operators

It is true that AI-based robots will be considered consumer products in some cases. But this hardly accounts for the entire scope of use of AI-based robots. With stronger cloud-computing abilities, robots may be operated, either physically or digitally, on an as-a-service model. In other words, AI-based robots may be used by operators for the purpose of offering services to the public (whether directly to consumers or to other businesses, which may or may not have end-users themselves).

In the physical context, think about a spa resort using robotic masseuses as alternative therapists at the spa. The clients obviously do not become owners of such robots and do not consume them as products; rather, the robots are used to provide services to the clients. In the digital context, imagine retaining the services of an AI-based DJ through an online service. The user opens an account through the service and the DJ-bot connects to the user's sound system and takes care of the music for the chosen event. Here too, the robot is used by an operator offering a service to its clients.

176. This idea of a "stricter" liability for manufacturers was a backlash to the negligence liability model that accompanied the industrial revolution. *See, e.g.*, Epstein, *supra* note 126, at 151–52; Keating, *supra* note 127, at 1290–91; Posner, *supra* note 108, at 29–32. Today, products liability doctrine mainly functions under a negligence model.

3. End-Users

Users are not commonly perceived as potential tortfeasors under tort doctrines related to product and consumer relationships. While users' actions may lead to the lack of liability of tortfeasors toward such users under contributory negligence doctrines,[177] limitations exceptions to products liability due to altering and modification of products,[178] or assumption of risk doctrines,[179] they are not commonly viewed as those who inflict harm or whose actions are being judicially reviewed as defendants.

Where AI-based robots are concerned, however, users may become tortfeasors themselves. This is due to the special nature of AI, which is necessarily based on interactions with external things (data, humans, or other machines).[180] Therefore, any interaction of an AI-based product with a user almost inevitably triggers reactions of the product that are unique and specific to the user. In such cases, interactions with users may lead to unforeseeable or unexpected product behavior that could inflict harm on others, thus potentially leading to user liability not only as a defense against manufacturer liability but also as a stand-alone or joint liability toward third parties.[181]

Some or all of the above stakeholders could have a part in a specific interaction with an AI-based robot, an interaction that may also inflict harm on any other stakeholders or third parties. The multiple-stakeholder problem will potentially make the liability analysis more complex.

VI. CONCLUSION

This Article attempted to review the special features of AI-based robots and the implications they may have on existing tort liability models. After suggesting a working definition of robots—non-human agents capable of demonstrating AI—and a working definition of AI—the program component of the robot that inherently causes it to act in a manner that is either inexplicable or unforeseeable to humans—it portrayed several important features of this phenomenon. AI-based robots may cause augmented harms in addition to physical injuries and damage to property, such as autonomy-based harms and privacy violations. One of the main characteristics unique to AI-based robots is the lack of personhood

177. RESTATEMENT (THIRD) OF TORTS: APPORTIONMENT OF LIAB. § 3 (AM. LAW INST. 2000) .

178. *See* RESTATEMENT (THIRD) OF TORTS: PRODS. LIAB. § 2 cmt. p, § 17 (AM. LAW INST. 1998).

179. RESTATEMENT (SECOND) OF TORTS § 496A (AM. LAW INST. 1965) ("A plaintiff who voluntarily assumes a risk of harm arising from the negligent or reckless conduct of the defendant cannot recover for such harm.").

180. Lemley and Casey referred to this phenomenon as a new type of harm or risk associated with robots—'misuse harms,' as they named it. *See* Lemley & Casey, *supra* note 17, at 1332–34.

181. This is not to argue that users cannot inflict harms on others using un-intelligent products. A consumer may well purchase a kitchen knife and stab her neighbor instead of cutting a salad. The uniqueness here is due to the fact that the harm will potentially be inflicted by the autonomous product, but the action, due to the lack of agency of the product, could be attributed (also) to the user. For a discussion framing human interaction intended to affect robots as hacking, see Ryan Calo et al., *Is Tricking a Robot Hacking?* (Univ. Wash. Tech. Policy Lab, Working Paper No. 2018–05, 2018). Another unique aspect of the AI case is that user interventions may cause a wider array of unforeseeable outcomes in comparison to un-intelligent products.

or agency of the entity behaving in a risky manner.[182] This is a significant challenge since tort law is rooted in the concept of liability for one's acts, save for specific doctrines of liability for the acts of another. The last, but definitely not least important, feature discussed was the inherent lack of foreseeability challenging basic principles in tort law, which requires foresight prior to imposing liability.

Following this general discussion, the Article proceeded to review the shortcomings of main tort law doctrines that may apply to harms inflicted by AI-based robots. Products liability doctrine seems to struggle with the lack of foreseeability characterizing AI-based robots, preventing a swift application of the design defect doctrine.[183] Abnormally dangerous activities cannot easily be attributed to AI-based robots for various reasons, but mainly because such robots may not at all be considered generally dangerous as a factual premise.[184] Even the general negligence doctrine falls short of fully capturing this phenomenon.[185] Two of negligence's four elements do not seem to fit with the concept of AI-based robots. The breach element is very difficult to establish due to the lack of foreseeability and explicability, a problem that also undermines the element of causation.[186]

Zooming out to a more general discussion on the appropriate liability regime for AI, we saw that none of the common regimes perfectly fits the challenges of AI-based robots. In the case of strict liability, neither designers, operators, nor end-users of robots may be best situated to assess the risks involved and the necessary preventive measures to be taken, taking the sting out of the main rationale for imposing strict liability. While a negligence regime (with comparative negligence) could generally apply in an optimal manner to this type of activity, it appears that determining the standard of due care would in many cases be very costly, mainly due to the foreseeability problem, thus undermining the efficiency of this type of liability without additional tools. Even insurance-based no-fault models cannot necessarily solve the problem due to significant difficulty in determining premiums and assessing risk expectancy, as well as to the cross-jurisdiction nature of AI-based robots.

Finally, the Article suggested imposing a predetermined level of care, using supplementary rules or quasi-safe-harbors, on different stakeholders, such as designers, distributors, operators, and end-users, that is better situated to employ them, thus creating a presumption of negligence. A monitoring duty could be imposed, based on technologies that do not require full understanding of the robots' behavior. If anomalies are detected, a duty to warn will follow, but this may not necessarily impose liability on the monitoring entity. Only the party best situated to take preventive measures after an anomaly has been detected will bear liability for expected harm. Another duty that could be imposed is including

182. *See supra* Section II.B.
183. *See supra* Section III.A.
184. *See supra* Section III.B.
185. *See supra* Section III.C.
186. *See supra* Section III.C.

emergency shut-down functions at the design stage. Ignoring this duty may lead to liability under the design defect doctrine, but abiding by it may shift the burden back to the designer under current negligence rules. Finally, ongoing support and patching duties, which follow insights from the monitoring duty, may be imposed on designers who may eventually have a duty to recall robots and patch them based on statistical inferences or case-specific behavior.[187] Failing to meet these predetermined standards would result in liability, while meeting them would revert the process to common negligence analysis with the aid of additional information generated in the process.

187. *See supra* Section V.C.