# THE REASONABLE ALGORITHM

*Karni Chagal-Feferkorn*†

*Abstract*

*Algorithmic decision-makers dominate many aspects of our lives. Beyond simply performing complex computational tasks, they often replace human discretion and even professional judgement. As sophisticated and accurate as they may be, autonomous algorithms may cause damage.*

*A car accident could involve both human drivers and driverless vehicles. Patients may receive an erroneous diagnosis or treatment recommendation from either a physician or a medical-algorithm. Yet because algorithms were traditionally considered "mere tools" in the hands of humans, the tort framework applying to them is significantly different than the framework applying to humans, potentially leading to anomalous results in cases where humans and algorithmic decision-makers could interchangeably cause damage.*[1]

*This Article discusses the disadvantages stemming from these anomalies and proposes to develop and apply a "reasonable algorithm" standard to non-human decision makers—similar to the "reasonable person" or "reasonable professional" standard that applies to human tortfeasors.*

*While the safety-promotion advantages of a similar notion have been elaborated on in the literature, the general concept of subjecting non-humans to a reasonableness analysis has not been addressed. Rather, current anecdotal references to applying a negligence or reasonableness standard to autonomous machines mainly discarded the entire concept, primarily because "algorithms*

† Fellow, Haifa Center for Law & Technology, University of Haifa Faculty of Law; PhD Candidate, University of Haifa Faculty of Law. I thank Niva Elkin-Koren for her endless support. I would also like to thank Yonathan Arbel, Ronen Avraham, Michal Gal, Nizan Geslevich-Packin, Elad Peled, Noa Mor-Golan, Dalit Ken-Dror, Maayan Perel-Filmar, Ronen Perry, Ariel Porat, Orna Rabinovich-Einy, Nadia Tzimerman, Asaf Yaakov, Keren Yalin-Mor as well as the participants of the Haifa Law & Technology Colloquium and the Minerva Center for the Rule of Law under Extreme Conditions Workshop for most helpful discussions and comments. This research was supported by the Center for Cyber Law & Policy (CCLP), established by the University of Haifa in collaboration with the Israeli National Cyber Bureau, the I-CORE Program of the Planning and Budgeting Committee and the Israel Science Foundation (1716/12) and by the Minerva Center for the Rule of Law under Extreme Conditions at the Faculty of Law and Department of Geography and Environmental Studies, University of Haifa, Israel and of the Israeli Ministry of Science, Technology and Space. Any mistakes or omissions remain the Author's.

1. *See generally* Sophia Duffy & Jamie P. Hopkins, *Sit, Stay, Drive: The Future of Autonomous Car Liability*, 16 SMU SCI. & TECH. L. REV. 101 (2013) (reviewing current legal frameworks that apply in the context of driverless cars); Jeffrey K. Gurney, *Imputing Driverhood: Applying a Reasonable Driver Standard to Accidents Caused by Autonomous Vehicles, in* ROBOT ETHICS 2.0. (chapter forthcoming 2016) (recommending that manufacturers of autonomous vehicles be treated as the drivers of their vehicles for purposes of assigning civil liability for harm caused by the vehicles' autonomous mode).

*are not persons."*[2] *This Article identifies and addresses the conceptual difficulties stemming from applying a "reasonableness" standard on non-humans, including the intuitive reluctance of subjecting non-humans to human standards; the question of whether there is any practical meaning of analysing the reasonableness of an algorithm separately from the reasonableness of its programmer; the potential legal implications of a finding that the algorithm "acted" reasonably or unreasonably; and whether such an analysis reconciles with the rationales behind tort law.*

*Other than identifying the various anomalies resulting from subjecting humans and non-humans conducting identical tasks to different tort frameworks, the Article's main contribution is, therefore, explaining why the challenges associated with applying a "reasonable standard" to algorithms are overcome.*

TABLE OF CONTENTS

I. Introduction ..... 113
II. Background ..... 116
A. Algorithms or Robots? ..... 116
B. Autonomous Algorithms ..... 117
C. The "Reasonable Person" Standard ..... 118
III. Applying a Reasonable Standard to Autonomous Algorithms ..... 121
A. Resolving the Anomalies ..... 122
1. Inequity Among Victims ..... 122
2. Procedural Inefficiency ..... 124
3. Economic Distortion ..... 125
4. Chilling Effects ..... 126
B. Technology Neutral Standard ..... 129
IV. Addressing the Conceptual Difficulties in Applying a "Reasonable Algorithm" Standard ..... 130
A. Intuitive Reluctance ..... 131
B. The "Homunculus Fallacy" ..... 132
1. Unpredictable Outcomes ..... 133
2. Time Lapse ..... 136
3. Different Standards of Reasonableness ..... 137
C. Legal Implications and Reconciliation with the Rationales Behind Tort Law ..... 139
1. Compensation ..... 142
2. Deterrence ..... 143
D. Would Superman be Subject to a Reasonableness Standard? ..... 144
V. Conclusion ..... 146

2. *See, e.g.*, Kyle Colonna, *Autonomous Cars and Tort Liability*, 4 W. RES. J.L. TECH & INT. 81, 102–04 (2012) (noting that applying the current negligence test to hardware or software is not practical because one cannot literally impute liability on a machine).

## I. INTRODUCTION

Algorithmic decision-makers have come to dominate various aspects of our lives. Many algorithms are characterized by machine-learning abilities, which enable them to make autonomous decisions[3] that replace judgement once reserved for humans.[4]

In the field of law, for example, "virtual attorneys" such as IBM's "Ross" have been deployed in law firms to conduct independent legal research,[5] algorithmic ODR mechanisms solve disputes online, often without any human facilitator,[6] and bail algorithms determine whether defendants awaiting trial may post bail to be released.[7] Attorneys and judges are not the only professionals relinquishing discretion to algorithms. This feature is also prevalent in many other professions and fields that require skills or expertise, where humans are either wholly replaced by an algorithm or a robot,[8] or surrender well-defined tasks to the "hands" (or "minds") of non-human decision makers. Physicians, for example, rely more and more on algorithms in order to diagnose medical conditions and select optimal treatments.[9] Forecasts predict that human drivers will gradually be replaced by driverless vehicles, while tax-returns experts as well as directors of companies are already being substituted by software.[10] In fact, even services provided by priests are no longer offered exclusively by human beings, but are provided by algorithms as well.[11]

---

3. The phrase "autonomy" or "autonomous" may have various meanings in the context of algorithms and machines. See further discussion below.

4. *See* Michal S. Gal, *Algorithmic Challenges to Autonomous Choice*, 23 MICH. TELECOMM. & TECH. L. REV. (2017) (discussing human choice and the implications of it being replaced by algorithms).

5. Anthony Sills, *Ross and Watson Tackle the Law*, IBM WATSON BLOG (Jan. 14, 2016), https://www.ibm.com/blogs/watson/2016/01/ross-and-watson-tackle-the-law; *Watson Takes the Stand*, ATLANTIC, http://www.theatlantic.com/sponsored/ibm-transformation-of-business/watson-takes-the-stand/283 (last visited Mar. 16, 2018).

6. Michael Legg, *The Future of Dispute Resolution: Online ADR and Online Courts*, AUSTRALASIAN DISP. RESOL. J. (forthcoming 2018), https://ssrn.com/abstract=2848097.

7. Tom Simonite, *How to Upgrade Judges with Machine Learning*, MIT TECH. REV. (Mar. 6, 2017), https://www.technologyreview.com/s/603763/how-to-upgrade-judges-with-machine-learning/.

8. For the sake of the discussion, the research would view both algorithms and robots interchangeably. For further discussion justifying this choice, please see below.

9. *See, e.g.*, Austin Frakt, *Your New Medical Team: Algorithms and Physicians*, N.Y. TIMES: THE UPSHOT (Dec. 7, 2015), https://www.nytimes.com/2015/12/08/upshot/your-new-medical-team-algorithms-and-physicians.html (noting that teams of physicians are helping to train Watson to apply humanity's huge store of cancer knowledge to the delivery of more personalized treatment); Vinod Khosla, *Technology Will Replace 80% of What Doctors Do*, FORTUNE (Dec. 4, 2012), http://fortune.com/2012/12/04/technology-will-replace-80-of-what-doctors-do/ (urging more technology involvement in healthcare because of the increasing amount of data and research).

10. *See, e.g.*, Richard Susskind & Daniel Susskind, *Technology Will Replace Many Doctors, Lawyers, and Other Professionals*, HARV. BUS. REV. (Oct. 11, 2016), https://hbr.org/2016/10/robots-will-replace-doctors-lawyers-and-other-professionals (expecting that within decades the traditional professions will be dismantled, leaving most professionals to be replaced by less-expert people, new types of experts, and high-performing systems); *see also* MCKINSEY GLOBAL INST., AUTOMATION POTENTIAL AND WAGES FOR US JOBS (2017), https://public.tableau.com/profile/mckinsey.analytics#!/vizhome/AutomationandUSjobs/Technicalpotentialforautomation (forecasting on the percentage of actions currently performed by human professionals to be replaced by atomization).

11. *Id.*; *see also* Jennifer M. Logg, *Theory of Machine: When Do People Rely on Algorithms*? (Harv. Bus. Sch. NOM Unit, Working Paper No. 17-086, 2017), https://ssrn.com/abstract=2941774 (discussing when people tend to prefer the advice of an algorithm and when they favor human input).

It is therefore not uncommon, and will likely become more and more frequent, that similar actions are performed interchangeably by humans and algorithms alike. Occasionally, these decision-makers can cause damage.[12] A pedestrian, for example, might be hit by a human driver, or might be similarly hit by a driverless car. A company might be adversely affected by a damaging decision reached by one of its human directors, or similarly affected by an identical decision made by an algorithmic director.[13] In the future, a patient might undergo a damaging surgery performed jointly by a human physician and a robo-doctor.[14]

Damages caused by human tortfeasors are judged under the well-established framework of negligence.[15] Under the negligence analysis, a wrongdoer is liable for damages if the four elements of negligence exist.[16] One of these elements is the "breach of a duty of care," determined by scrutiny of comparable decisions a "reasonable person" would reach under similar circumstances.[17] When a wrongdoer acted "reasonably," therefore, the wrongdoer, and other parties which might be vicariously involved, would be found not liable.[18] For example, if a human surgeon caused damage, she and the hospital that employed her would be free from liability if the doctor was found to have acted reasonably.[19]

No similar "reasonableness analysis" currently exists, however, for identical damages caused in an identical matter to identical victims, when the wrongdoer is not human.[20]

Rather, damaging algorithms have generally been treated as "products" or "tools" in the hands of their manufacturers or users, and are therefore subject to legal frameworks such as product liability or direct negligence of the humans

12. *See, e.g.*, Ashley Halsey III, *Transportation When Driverless Cars Crash, Who Gets the Blame and Pays the Damages?*, WASH. POST (Feb. 25, 2017), https://www.washingtonpost.com/local/trafficandcommuting/when-driverless-cars-crash-who-gets-the-blame-and-pays-the-damages/2017/02/25/3909d946-f97a-11e6-9845-576c69081518_story.html (noting that while computer-driven cars are expected to reduce crashes dramatically, nobody in the field thinks collisions will become a thing of the past).

13. *Algorithm Appointed Board Director*, BBC NEWS (May 16, 2014), http://www.bbc.com/news/technology-27426942.

14. Michael MacRae, *The Robo-Doctor Will See You Now*, ASME (May 2012), https://www.asme.org/engineering-topics/articles/robotics/robo-doctor-will-see-you-now.

15. *See generally* RESTATEMENT (SECOND) OF TORTS § 281 (AM. LAW INST. 1965); WILLIAM L. PROSSER & WERDNER P. KEETON, PROSSER AND KEETON ON THE LAW OF TORTS (W.P. Keeton 5th ed., 1984) (describing tort law).

16. *Id.*

17. *See, e.g.*, Alan D. Miller & Ronen Perry, *The Reasonable Person*, 87 N.Y.U. L. REV. (2012) (explaining how in the case of professional decisions, "reasonableness" is evaluated in comparison to the decisions expected from "reasonable professionals").

18. RESTATEMENT (SECOND) OF TORTS § 429 (AM. LAW INST. 1965).

19. *See* Howard Levin, *Hospital Vicarious Liability for Negligence by Independent Contractor Physicians: A New Rule for New Times*, 2005 U. ILL. L. REV. 1291 (Oct. 2005) (noting that in the past, hospitals' liability for damages caused by physicians was very limited, owed to the "independent contractor" doctrine); Joseph Magnet, *Ostensible Agency in American Hospital Law: Does Canada Need It?*, CANADIAN CASES L. TORTS 187 (1980) (noting that legislation updates, however, render hospitals liable for such damages caused by their physicians nevertheless).

20. *See, e.g.*, Duffy & Hopkins, *supra* note 1 (reviewing current legal frameworks that apply in the context of driverless cars).

involved.[21] Different features of algorithmic decision-makers, such as their improving self-learning abilities and the lack of foreseeability of their choices, has raised much debate on the tort legal framework that ought to apply to them, and the identity of the parties that should assume liability for their damages.[22] The European Parliament, for example, has issued a draft report explaining that autonomous robots can no longer be considered tools in the hands of other actors, suggesting to award autonomous robots with an independent legal status of "electronic persons," which would entail the ability of said "electronic persons" to pay damages themselves.[23]

The current Article will, however, remain in a more traditional territory where algorithms do not bear liability for their own actions. Rather, the Article assumes that, at least in the near future, liability for algorithmically-caused damages would continue to rest with the natural persons or legal persons involved. The Article suggests, however, that the framework for determining whether said liability exists would be more similar to the one that applies to damages caused by humans. In more detail, the Article assumes that while liability for damages would be imposed on humans or legal entities, it is the action (or decision) of the algorithm itself[24] that must be scrutinized for "reasonableness" rather than the decisions of the humans involved.[25]

The limited literature on applying reasonableness (or, more generally, negligence) standards on the technology itself has for the most part rejected this framework without much discussion, claiming primarily that "algorithms are not persons" and therefore cannot be subject to, or do not warrant, an independent analysis of reasonableness.[26] For instance, because a software is not a "person"

21. *See, e.g.*, *id.* Another algorithmic example involving both frameworks of product liability and direct negligence by the humans involved could be found in aviation accidents attributed to autopilots. See, for example, the 2013 Asiana-Air crash in San-Francisco, where the underlying legal actions consisted of product liability claims raised against Boeing, the manufacturer of the auto throttle that allegedly failed, as well as negligence claims raised against the airline itself. Matt Hamilton, *Asiana Crash: 72 Passengers Settle Lawsuits Against Airline*, L.A. TIMES (Mar. 3, 2013), http://www.latimes.com/local/lanow/la-me-ln-asiana-airlines-settle-lawsuits-20150303-story.html.

22. In the case of driverless vehicles, for example, the literature discusses whether it is the user of the car or its manufacturer that ought to be liable for car accidents the vehicle was involved in. Duffy & Hopkins, *supra* note 1.

23. "Whereas the more autonomous robots are, the less they can be considered simple tools in the hands of other actors (such as the manufacturer, the owner, the user, etc.); whereas this, in turn, makes the ordinary rules on liability insufficient and calls for new rules which focus on how a machine can be held—partly or entirely—responsible for its acts or omissions . . . ." Eur. Parl. Draft Rep. on Civ. L. Rules on Robotics, 2015/2103 (INL), at 6 (Jan. 1, 2017).

24. Providing specific guidelines to differentiate algorithms that warrant an independent analysis of reasonableness from those that may still be considered a mere "tool" or "product" would be the matter of a separate article. This Article therefore uses the general terms "algorithms," "learning algorithms," or "autonomous algorithms" in reference to those that should be subject to the reasonableness test. See further discussion on the meaning of "autonomy" in the algorithmic context below.

25. If, for example, a robo-doctor acted reasonably, then a possible implication could be that the robo-doctor itself as well as the hospital that owned or used it, and possibly the manufacturer who created it would not be subject to liability (similar to the implication of finding that a flash and blood physician acted reasonably, as discussed above). Part III discusses this further and reviews other possible implications for a finding of reasonableness or unreasonableness on the part of the algorithm.

26. Kyle Colonna, *Autonomous Cars and Tort Liability*, 4 CASE W. RES. J. L. TECH & INTERNET 81, 102–04 (2012). Alternatively, scholars have suggested applying a reasonableness test to the technology itself in very specific contexts, such as autonomous cars' liability, where the discussion focused mainly on why said standard

and cannot pay damages if found liable,[27] or because algorithms act as they are programmed to act and thus do not warrant an independent analysis of reasonableness.[28] The Article will address said concerns and others, and explain how they are overcome. To do so it will, among other things, explain why algorithms' "choices" might be deemed reasonable while at the same time the choices of their programmers deemed unreasonable, and vice versa, and why the main rationales behind tort law could be met even when a "reasonableness standard" is applied to the algorithm itself.

Chapter II provides general background on learning algorithms and on the concept of "reasonableness." Chapter III then discusses the anomalies and negative effects resulting from applying different legal frameworks to humans and non-humans engaging in a similar decision-making activity. Chapter IV focuses on the suggestion to develop and apply a "reasonable algorithm" standard that would eliminate or minimize the anomalies discussed in Chapter III, and explains why the conceptual difficulties associated with said notion are overcome.

## II. BACKGROUND

### *A. Algorithms or Robots?*

While an interesting research question would pertain to 'when would a person relying on an algorithm's recommendation be deemed reasonable,' this Article deliberately does not focus on the reasonableness of human beings (be it the person who relied on the machine or the person who programmed it). Rather, it chooses to focus on the more controversial question of whether it is sensible to apply a reasonableness standard to the algorithm itself. An obvious preliminary question that comes to mind, therefore, is whether we should only be focusing on robots, which, unlike algorithms, have a physical embodiment and can therefore carry out damaging actions on their own, without the involvement of humans.

Indeed, when this Article refers to "algorithms", it does not refer to written decision-trees. Rather, "algorithms" in this Article mean algorithms that are computerized. Such algorithms are capable of causing damage without any human intervention or without any physical embodiment other than the hardware of the computer, for example by giving problematic trading orders, or as part of the Internet of Things (IOT) revolution where machines give orders to each other without human involvement.[29] The Article therefore follows Balkin's

would entail less litigation costs than the alternative framework of product liability; *see, e.g.*, JEFFREY K. GURNEY, ROBOT ETHICS 2.0. 51–65 (Patrick Lin et al. eds., 2016); or offered it in the context of encouraging technological advancement and improved safety: *see generally* Ryan Abbot, *The Reasonable Computer: Disrupting the Paradigm of Tort Liability*, 86 GEO. WASH. L. REV. 1 (2017).

27. Colonna, *supra* note 26.

28. Jack M. Balkin, *The Three Laws of Robotics in the Age of Big Data*, 78 OHIO ST. L.J. (Sep. 10, 2017).

29. *See* Michal S. Gal & Niva Elkin-Koren, *Algorithmic Consumers*, 30 HARV. J.L. & TECH. (2017) (discussing the future of e-commerce and how it would be ruled by algorithmic agents bypassing human decision).

classifications, which treat robots and algorithms alike, both being similar members of the "algorithmic society."[30]

Granted, robots' physical embodiment may have different implications within the tort law framework. For example, people tend to react to robots with human-like appearance similarly to how they react to real persons (including defending them in the battlefield or avoiding actions they would feel uncomfortable committing in the presence of a real person).[31] Said difference may have relevance in the context of tort litigation (for example, if people would tend to avoid lawsuits against "cute" damaging robots, but will not avoid them when the tortfeasor is an amorphous algorithm). Most of the arguments discussed in this Article, however, focus on the machine's decision-making process. In that context, we are interested in the "mind" behind the decision, whether or not it has a physical "body."

### *B. Autonomous Algorithms*

The abilities of algorithms are advancing, such that they perform complex actions and make intricate decisions that require abilities that far exceed mere computations.[32] For example, different algorithms assist medical staff in diagnosing medical conditions, matching the optimal treatment to each patient, and even physically performing certain medical procedures.[33] In finance, algorithms are used for assessing credit risks and mortgage risks, pricing complex insurance products, stocks ranking, or in general, creating financial forecasts.[34] Deference to algorithmic decision-making is common in the professional context, but also in the context of transportation, culture, consumption, and tourism,[35] to name just a few fields.

A foremost factor leading to a dramatic increase in algorithms' ability to "take over" actions once reserved for humans, is their "self-learning" ability, known as "machine learning."[36] The advancement of technology and the prevalence of enormous amounts of data allow algorithms to learn from existing information and implement the conclusions in future sets of data.[37] Learning can be "supervised," where algorithms train on a portion of the data and are

30. Balkin, *supra* note 28.

31. Ryan Calo, *Robotics and the Lessons of Cyberlaw*, 103 CALIF. L. REV. 513, 550–58 (2015).

32. Gal & Elkin-Koren, *supra* note 29.

33. Frakt, *supra* note 9; Khosla, *supra* note 9.

34. Amir E. Khandani et al., *Consumer Credit Risk Models Via Machine-Learning Algorithms*, 34 J. BANKR. & FIN. 2767 (2010); Justin Sirignano et al., *Deep Learning for Mortgage Risk*, (Working Paper, 2018), https://arxiv.org/pdf/1607.02470.pdf; Anna Bacinello et al., *Regression-based Algorithms for Life Insurance Contracts with Surrender Guarantees*, 10 QUANTITATIVE FIN. 9 (Oct. 22, 2009), http://ssrn.com/abstract=1028325; Ying Becker et al., *An Empirical Study of Multi-Objective Algorithms for Stock Ranking*, SSRN (Jun. 26, 2007), http://ssrn.com/abstract=996484; Babak D. Mahdavi, *Machine Learning Methods for Financial Forecasting: Application to the S&P 500*, SSRN (2006) (on file with Author).

35. Marcus du Sautoy, *How Do Algorithms Run My Life?*, BBC NEWS (Sept. 24, 2015), http://www.bbc.co.uk/guides/z3sg9qt; Sean O'Neill, *Startup Pitch: IBM's Watson Powers New Travel Advice Tool WayBlazer*, TNOOZ (Oct. 7, 2014), https://www.tnooz.com/article/startup-pitch-wayblazer-aims-travel-insights-service/.

36. Amir Gandomi & Murtaza Haider, *Beyond the Hype: Big Data Concepts, Methods and Analytics*, 35 INT'L. J. INFO. MGMT 137, 144 (2015).

37. *Id.*

given correct answers to the training tasks, so that they may create, on their own, a model for solving future tasks pertaining to similar data.[38] The learning stage may also be "unsupervised" such that the algorithm is not "fed" any answers but is "free" to decipher patterns in the data that may indicate the right answer.[39] Though the degree of freedom for the algorithm to make its own choices could be greater under unsupervised learning, both methods, as will be explained later, involve lack of foreseeability by their human developer.[40]

Naturally, progress in algorithmic capability to mimic human skills, among them self-learning, has increased their "autonomy" level.[41] The phrase "autonomous algorithm" or "autonomous decision-maker" is widely used, but its meaning varies.[42] By "autonomy," some mean a trait of the algorithm (or machine) itself: its ability to "understand" its actions or their consequences, or to "teach itself" how to perform certain tasks.[43] For others, "autonomy" is the level of authorization the algorithm (machine) has to act on its own, without a human's permission.[44] Others differentiate a futuristic "substantial autonomy," where a system possesses its own self-awareness and freedom of choice, from a "technical autonomy," where the system is free to choose between pre-programmed options.[45] In fact, it is sometimes argued that any attempt to define "autonomy" will inevitably be based on controversial assumptions.[46]

A future article may delve into the specific types of "autonomous" algorithms that deserve their own separate analyses of reasonableness. This Article's scope is limited to those "autonomous algorithms" that have self-learning abilities and can often yield results not foreseeable by their programmers.

### *C. The "Reasonable Person" Standard*

A "tortious act" is a wrong in which a tortfeasor has caused a victim harm.[47] The framework of tort law is dedicated to determining under which

38. *Id.*

39. Avigdor Gal, *It's A Feature, Not A Bug: On Learning Algorithms and What They Teach Us*, OECD (June 7, 2017), https://one.oecd.org/document/DAF/COMP/WD(2017)50/en/pdf; Harry Surden, *Machine Learning and the Law*, 89 WASH. L. REV. 87 (2014).

40. THOMAS B. SHERIDAN & WILLIAM L. VERPLANK, HUMAN AND COMPUTER CONTROL OF UNDERSEA TELEOPERATORS (1978), http://www.dtic.mil/dtic/tr/fulltext/u2/a057655.pdf.

41. *See Getting Machines to Mimic Intuition*, SIEMENS (Apr. 20, 2016), https://www.siemens.com/innovation/en/home/pictures-of-the-future/digitalization-and-software/autonomous-systems-machine-learning.html ("The ability to learn is a precondition for autonomy."); *see* Harry Surden, *Machine Learning and Law*, 89 WASH. L. REV. 87, 95 (2014) (explaining that "researchers have successfully used machine learning to automate a variety of sophisticated tasks that were previously presumed to require human cognition").

42. *See, e.g.*, SHERIDAN & VERPLANK, *supra* note 40, at 1–3, (discussing different types of autonomous algorithms).

43. *Id.* at 1.

44. *Id.* at 1–3.

45. Eliav Lieblich & Eyal Benvenisti, *The Obligation to Exercise Discretion in Warfare: Why Autonomous Weapon Systems are Unlawful*, *in* AUTONOMOUS WEAPONS SYSTEMS: LAW, ETHICS, POLICY 244 (Nehal Bhuta et al. eds., Cambridge Univ. Press 2016), https://ssrn.com/abstract=2479808.

46. Noel Sharkey, *Staying in the Loop: Human Supervisory Control of Weapons*, *in* AUTONOMOUS WEAPONS SYSTEMS: LAW, ETHICS, POLICY (Nehal Bhuta et al. eds., Cambridge Univ. Press 2016).

47. JOHN C. P. GOLDBERG & BENJAMIN C. ZIPURSKY, THE OXFORD INTRODUCTIONS TO U.S. LAW: TORTS (2010).

circumstances a wrong-doer must pay the victim compensation.[48] The decision must strike a balance between deterring potential tortfeasors from committing torts and allowing victims to recuperate, while also enabling people and entities to perform beneficial and desirable deeds without paralyzing concern about tort liability.[49] The point where the balance is reached is determined by many parameters, but in general depends on the specific rationales behind tort law that society wishes to achieve and the emphasis it lays on each.

A leading rationale behind tort law is the forward-looking concept of *deterrence*: shaping the legal system such that potential tortfeasors take precautions that will prevent the execution of future torts. Under "optimal deterrence," a tortfeasor is liable for the tort only if the cost of the harm she has caused exceeds the cost of precautions she could have taken to prevent the tort.[50]

A second dominant rationale behind tort law is the backward-looking concept of *compensation*.[51] This stems from a "rights-based" principle of corrective justice, where the tortfeasor is required to correct the wrong she has committed, based on justice and fairness considerations.[52]

To establish negligence, four elements must be proven: the existence of a duty of care, a breach of that duty, causation between the breach of duty of care and the damage and the existence of damage.[53] A person breaches her duty of care if she does not adhere to the standard of reasonable care when carrying out actions that might foreseeably harm others.[54] To determine whether "reasonable care" was demonstrated or not, courts resort to the "reasonable person" standard, which asks what a reasonable person would have done under similar circumstances and possessing the same state of knowledge.[55]

Alas, "reasonableness" is in the eyes of the beholder. The answer to whether a reasonable person would or would not have done something under certain circumstances depends on different values of reasonableness poured into

48. *See* SAUL LEVMORE & CATHERINE M. SHARKEY, FOUNDATIONS OF TORT LAW (2012) ("[T]he academic discussion of tort law in the United States focuses primarily on deterrence and corrective justice, largely because both of these theoretical approaches seek to explain why tort law renders compensation to injured parties.").

49. *The Functions and Goals of Tort Law*, *in* TORT LAW AND PRACTICE (Dominick Vetri et al. 2016); Benjamin Shmueli, *Legal Pluralism in Tort Law Theory: Balancing Instrumental Theories and Corrective Justice*, 48 U. MICH. J.L. REFORM 745, 745–46 (2015).

50. Richard A. Posner, *The Value of Wealth: A Comment on Dworkin and Kronman*, 9 J. LEG. STUD. 243, 244 (1980); John C. P. Goldberg, *Twentieth Century Tort Theory*, 91 GEO. L.J. 513, 552 (2002).

51. *See* ERNEST WEINRIB, THE IDEA OF PRIVATE LAW 5 (1995) ("[C]ompensation and deterrence [are] the two standard goals ascribed to tort law."); Shmueli, *supra* note 49, at 752 (stating that some theoreticians consider compensation as an independent and even predominant aim of tort law).

52. WEINRIB, *supra* note 51, at 132. The compensation rationale might also be compatible with the rationale of efficiency and optimal deterrence. For instance, without compensation, the victim might endure additional costs that society would have to bear. *See generally* Mark Geistfeld, *Efficiency, Fairness, and the Economic Analysis of Tort Law* (N.Y.U. Public Law & Legal Theory Research Paper Series Working Paper No. 09-26, 2009), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1396691; *see generally* THEORETICAL FOUNDATIONS OF LAW AND ECONOMICS (Mark D. White et al. eds., 2008).

53. RESTATEMENT (SECOND) OF TORTS, *supra* note 15, at § 281; PROSSER & KEETON, *supra* note 15.

54. RESTATEMENT (THIRD) OF TORTS: LIABILITY FOR PHYSICAL AND EMOTIONAL HARM § 3 (AM. LAW. INST., 2010); Benjamin C. Zipursky, *Foreseeability in Breach, Duty, and Proximate Cause*, 55 WAKE FOREST L. REV. 1247, 1249–50 (2009).

55. Miller & Perry, *supra* note 17, at 325.

this very general and subjective term.[56] First, reasonableness can be measured per a *positive standard*, which compares the tortfeasor's behaviour with the behaviour of others: a tortfeasor did not breach her duty of care if she acted as others would have.[57] Alternatively, courts may choose to apply a *normative standard*, which asks what a reasonable person should have done and aims to direct potential tortfeasors' behaviour according to desired values.[58] Moreover, another dimension of complexity is invoked regarding whether the reasonable standard should be objective or subjective.[59]

In general, the reasonable standard is objective, when at the same time the specific tortfeasor's state of knowledge is considered.[60] For example, with respect to professionals, those acting within the scope of their profession would generally be subject to the elevated standard of "the reasonable professional" rather than the "reasonable person."[61] In addition, when judging the alleged negligence of professionals, courts usually resort to normative rather than positive standards of reasonableness (with the exception of medical malpractice which is generally analysed on the basis of a positive "custom-based standard of

56. *Id*. at 379–80.

57. The reference point of "others" is complex within itself, and is given different meanings over time and in different courts. For example, a court might compare the reasonableness of a tortfeasor with that of a "common" or "ordinary" reasonable person, or with that of a "prudent" or "ideal" reasonable person. Even if deciding that the reference point is indeed an "ordinary" person, "ordinary" could be based on what the majority of people would have done, or on what the "average person" would have done, or perhaps on what the "median person" would have done (complicating this standard even further is what "majority," "average," or "median" are measured by). Henry T. Terry, *Negligence*, 29 HARV. L. REV. 40 (1915); Miller & Perry, *supra* note 17, at 370.

58. This standard is no less vague than the positive standard, as what a reasonable person should have done naturally depends on the specific values that the legal system wishes to promote. An example of a normative value is the Kantian idea of equal freedom to act in ways that coexist with the freedom of others. By this approach, "reasonable care" has to be shaped to reconcile one's liberty to act (given that any action might entail a possible risk) with one's freedom not to be harmed by another's actions. Miller & Perry, *supra* note 17, at 351; IMMANUEL KANT, THE METAPHYSICS OF MORALS § 237 (Mary Gregor ed., Mary Gregor trans., 1991). Another well-known example of a normative standard of reasonableness is economic efficiency (welfare maximization), formulated by Judge Learned Hand. According to the Hand formula, liability has to be imposed when the cost of taking precautions to prevent the damage is less than the damage expectancy. Consequently, determining that a person acted unreasonably when failing to take cost-effective measures results in internalizing the externalities of inefficient actions by potential tortfeasors. Thus, optimal deterrence, which as noted is one of the main rationales behind tort law, is obtained. Though the two approaches might yield very similar results when one analyses specific cases of "reasonableness", they differ in that far-fetched risks will be deemed reasonable by the Kantians, regardless of the precautions taken, while "real" foreseeable risks will be deemed unreasonable, even if effective precautions were taken. This is because under the Kantian approach "[a] far-fetched risk is the kind that every person is prepared to endure, knowing that all human activity involves such risks and that trying to eliminate them would disable action. Conversely, no one is willing to be exposed to real risks. Since protection of all humans must be equal, every person must waive the possibility of exposing others to risks of that magnitude." Miller & Perry, *supra* note 17, at 351–52.

59. Or, in other words, whether the tortfeasor's specific attributes and characteristics be considered. While a fully *subjective reasonableness standard* would be meaningless, a fully *objective reasonableness standard* would ignore relevant data that could be critical for assessing the tortfeasor's action (such as a physical disability that caused the tortfeasor, for example, to misinterpret what she had seen). *See* Victoria Nourse, *After the Reasonable Man: Getting over the Subjectivity/Objectivity Question*, 11 NEW CRIM. L. REV. 33 (2008) (discussing both subjective and objective reasonableness approaches and suggesting that the correct approach is the hybrid standard that incorporates both approaches).

60. *See id*. at 36 ("[A] majority of jurisdictions adopt a standard that is both objective and subjective.").

61. RESTATEMENT (SECOND) OF TORTS §§ 289, 299A (AM. LAW INST. 1965) (highlighting how the higher the degree of professionalism characterizing the tortfeasor, the higher the standard of reasonableness she must adhere to); ARTHUR BEST & DAVID W. BARNES, BASIC TORT LAW: CASES, STATUTES AND PROBLEMS (2007).

care").[62] These and other variations in the interpretation of "reasonableness" would undoubtedly be very significant at the later stage of deciding which specific content to pour into the actual "reasonable algorithm" standard. For the purpose of this Article, suffice it to remember that "reasonableness" is a broad standard that could be interpreted in various ways.

Having reviewed the basic information pertaining to autonomous algorithms on the one hand, and the general standard of reasonableness on the other, Part III will address the disadvantages that arise from applying different tort frameworks on similar actions performed by autonomous algorithms and humans, as well as the advantages of applying a "reasonableness" standard on both.

## III. APPLYING A REASONABLE STANDARD TO AUTONOMOUS ALGORITHMS

To support the application of a "reasonable algorithm" standard, a logical step would be to examine the advantages of applying such a standard. For a fruitful discussion, however, we must compare the application of a reasonableness standard to algorithms with other specific alternatives (for example, product liability, no fault insurance schemes, etc.), and examine in which sense the former is better than the latter. A reasonableness standard, might lead to lower trial costs under the product liability framework, and thus give potential victims greater access to justice.[63] In comparison to strict liability (which applies in certain product liability cases),[64] the more lenient reasonableness standard may encourage the usage of machines and thus promote innovation and improve safety.[65] Likewise, a reasonableness standard may create more effective deterrence than a rule of no-fault liability.[66] A comprehensive analysis will obviously require specific content to be poured into the 'reasonable algorithm' standard first. Since the guidelines for development of the content itself is beyond the scope of this Article, the Article does not purport to compare this standard with other mechanisms for imposing liability on algorithms. Instead, it focuses on arguments pertaining to the "reasonable

62. John W. Ely et al., *Determining the Standard of Care in Medical Malpractice: The Physician's Perspective*, 37 WAKE FOREST L. REV. 861 (2002); James A. Henderson, Jr. & John A. Siliciano, *Universal Health Care and the Continued Reliance on Custom in Determining Medical Malpractice*, 79 CORNELL L. REV. 1382 (1994); *see* James Gibson, *Doctrinal Feedback and (Un)Reasonable Care*, 94 VA. L. Rev. 1641 (2008) (arguing, thus, medical choices that are in line with common protocols or clinical guidelines will usually be protected from tort liability); *see* Philip G. Peters, *The Quiet Demise of Deference to Custom: Malpractice Law at the Millennium*,163 WASH. & LEE L. REV. 57 (2000) (describing by this standard, physicians will usually be found to have acted reasonably when complying with customary practice); *cf.* Glen O. Robinson, *Rethinking the Allocation of Medical Malpractice Risks Between Patients and Providers*, 49 L. & CONT. PROB. 1, 173 (1986).

63. Colonna, *supra* note 26; Jeffrey K. Gurney, *Sue My Car Not Me: Products Liability and Accidents Involving Autonomous Vehicles*, 2013 U. ILL. J.L. TECH. & POL'Y 247 (2013).

64. *See* GOLDBERG & ZIPURSKY, *supra* note 47 (arguing that in most states, strict liability is applied to "manufacture defects," where the product was not properly manufactured, due to diverting from the product's assembly specifications or to using non-appropriate materials).

65. Abbot, *supra* note 26.

66. *See, e.g.*, Alan Marco & Casey Salvietti, *What Does Tort Law Deter? Precaution and Activity Levels in No-Fault Automobile Insurance*, (2nd Ann. Conf. on Empirical Legal Stud. Paper, Nov. 11 2007), https://ssrn.com/abstract=998741 (discussing the effect of no-fault automobile insurance regime on deterrence).

algorithm" standard on its own, especially those addressing the uniformity of the tort framework that would apply to algorithmic and human tortfeasors alike.

### *A. Resolving the Anomalies*

Subjecting humans and algorithms to different tort frameworks for performing the same actions may result in certain anomalies. The following describes these anomalies and their potential disadvantages, thereby showing why applying the same type of legal analysis to both human and algorithmic damages could be advantageous.

A preliminary comment is that the specific expectations we would have from the "reasonable algorithm" might very well differ from the expectations a "reasonable person" or a "reasonable professional" is measured against. Intuitively, algorithms' superior abilities may lead to an elevated level of reasonableness which would be required of them[67]—although a more complex standard could possibly be adopted, based also on algorithms' weaknesses discussed below.[68] Some of the following arguments will therefore depend on the exact content poured into the reasonable algorithm standard, and on the size of the gap between that content and the content applying in the context of human actions. The discussion will nevertheless assume that, at least for the most part, applying a "reasonableness" standard on both types of decision makers[69] would result in much greater similarities than when humans are subject to one framework while algorithms are subject to a completely different one.

#### *1. Inequity Among Victims*

As discussed above, one of the main rationales behind tort law is compensating the victim.[70] From this perspective, differential treatment raises important challenges, because it might expose victims to lower payouts and a higher risk in comparison to other similar victims.[71]

For example, imagine pedestrian A crossing the road at a certain point and getting hit by a car. Under a negligence cause of action, A would file suit against the driver, and be awarded compensation if the driver had acted unreasonably. Now imagine pedestrian B, A's twin, hit at the same place under the same circumstances, and suffering the same damage, except the hitter was a driverless car. The type of legal action B would have to pursue would, in its complexity, costs, effort and time required, chances of receiving compensation, and its amount, likely be entirely different from those associated with A's legal action, as long as the legal framework applying to damage caused by driverless cars differs from that of negligence of the hitter. If, for example, damages caused by

67. GURNEY, *supra* note 26; David Vladeck, *Machines Without Principals: Liability Rules and Artificial Intelligence*, 89 WASH. L. REV. 117, 135–36 (2014).

68. *See* Karni Chagal-Feferkorn, *Who Are You, the Reasonable Algorithm*? (forthcoming 2018) (discussing the content of the standard of reasonableness ought to be developed for algorithms).

69. GURNEY, *supra* note 26; Vladeck, *supra* note 67.

70. *See generally* Marco & Salvietti, *supra* note 66 (discussing the overall goals of tort laws).

71. *See* John C. P. Goldberg & Benjamin C. Zipursky, *Tort Law and Moral Luck*, 92 CORNELL L. REV. 1123 (2007) (describing the different ways luck affects tort law).

the driverless car were subject to strict liability under product liability rules, then B would have to show that the vehicle was defective, without addressing questions of culpability by any party.[72] A, on the other hand, would need to show that a human driver acted unreasonably—a burden of proof that focuses on completely different issues and will likely entail different legal costs and probabilities of success.[73]

Treating victims in the same class differently may infringe upon notions of horizontal equity. Under this principle, like-cases should be treated alike.[74] In other words, justice requires that victims should be treated similarly by the legal system regardless of the identity of their injurer.[75] Allowing one victim to recover quickly and easily while subjecting another victim of similar damage to a lengthy, costly, and uncertain procedure interferes with this ideal.

Granted, horizontal equity is more of an ideal than a practical standard, with or without the involvement of algorithmic decision-makers;[76] sheer luck and circumstantial reasons frequently have a crucial effect on victims' redress.[77] Pedestrian C, for example, might be hit by a very wealthy driver and be offered an immediate compensation sufficient to cover her damages and more. Pedestrian D, on the other hand, might suffer the exact same injury under equal circumstances, but spend months in legal proceedings against the driver who hit her, only to discover in the end that the driver's resources do not suffice for any compensation.[78] It would be nothing but D's misfortune to have been hit by that specific driver, and not a more affluent one, that had led to such a different outcome than in C's case. In addition, even when both drivers are capable of compensating the victim, horizontal equity is rarely implemented in full: the difference between judicial instances, between the capabilities of the attorneys involved and between states' laws will often lead to deviations from a truly equal treatment of similar cases.[79] However, said difference presents horizontal inequity at its most crystalized form: even if both tortfeasors have sufficient resources to compensate the victims, and even if the cases were judged at the same instance by the same judges and were argued by the same lawyers, the inherent differences between the reasonableness standard that would apply to the human driver and the wholly different tort framework that might apply to the driverless car would render the two cases, their costs and their outcomes, very different.[80] In addition, while deviations in treatment stemming from misfortune or unequal circumstances are a matter of pragmatic necessity, applying a

72. RESTATEMENT (SECOND) OF TORTS § 402A (AM. LAW INST. 1965); GOLDBERG & ZIPURSKY, *supra* note 47 (describing how product liability covers three different types of "defects," among them "design defects").

73. See discussion on the differences between the legal questions raised under each of these frameworks below.

74. JOSEPH W. DOHERTY ET AL., CONFIDENTIALITY, TRANSPARENCY, AND THE U.S. CIVIL JUSTICE SYSTEM 119–24 (2012).

75. *Id.*

76. *Id.*

77. Goldberg & Zipursky, *supra* note 71.

78. *Id.* (discussing the different ways luck affects tort law).

79. Stephen D. Sugarman, *A Comparative Law Look on at Pain and Suffering Awards*, 55 DEPAUL L. REV. 399, 413 (2006).

80. GOLDBERG & ZIPURSKY, *supra* note 47; Colonna, *supra* note 26; Gurney, *supra* note 63.

different legal framework on similar victims is a deliberate choice to subject them to the arbitrariness of sheer luck. Though said choice might be justified due to policy considerations, on its own the lack of equal treatment of equal victims is infringing upon principles of fairness. Applying a reasonableness standard to damages caused by algorithms (depending, of course, on the specific 'reasonableness' test developed for algorithms and how similar they were to those for a person) would generally create more unity in the type of procedure and burden of proof that a victim must meet, and in the expected costs and outcomes, and increase the horizontal equity among victims.

### *2. Procedural Inefficiency*

A different type of anomaly would arise when the victim suffered damage from an algorithm and a human being as joint tortfeasors. Imagine a patient treated by two different physicians: for example, a woman in labour being treated by two different doctors on duty, who consult with each other and jointly decide to avoid caesarean section—a decision that ultimately caused harm. Now assume that one of the two decision makers was not human but rather an algorithm or a robot, which are not subject to the reasonableness standard. Naturally, the patient's legal proceeding against the two tortfeasors would become much more complex than in the case both tortfeasors were subject to a similar tort framework.[81]

Indeed, it is not uncommon that a single legal action invokes several causes of action governed by different legal frameworks.[82] When one of the joint tortfeasors is a machine, however, the lack of a unified framework might increase the time and costs associated with the legal proceeding's many folds.[83] This is because the different causes of action against the human tortfeasor subject to the negligence standard versus the algorithmic wrongdoer subject to a different legal framework would raise different legal questions, but would also probably require very dissimilar sets of arguments, of evidence, and of expert opinions.[84] For example, if the victim invokes a claim of a "design defect" under product liability[85] against the algorithmic doctor, this would require that she shows that a feasible safer alternative design of the algorithm could have been used or, in certain states, that the risk posed by the product exceeds the expectations of an ordinary consumer.[86] Though the former test focuses on the algorithm itself (whether it could have been safer or not), the analysis would be

81. *See generally* Patrick F. Hubbard, *'Sophisticated Robots': Balancing Liability, Regulation and Innovation*, 66 FLA. L. REV. 1803, 1843 (2014).

82. *See* Hamilton, *supra* note 21 (discussing aviation accidents where both product liability and negligence claims were raised and different causes of actions were brought against a single defendant).

83. Hubbard, *supra* note 81, at 1811.

84. UGO PAGALLO, THE LAWS OF ROBOTS: CRIMES, CONTRACTS AND TORTS 135 (Springer Science & Business Media Dordrecht 2013).

85. Allegations of defects of damaging algorithms are likely to fall under "design defects" rather than other types of product defects. Hubbard, *supra* note 81, at 1854 (stating that allegations of design defect refers to a product that could have been designed in a safer manner that is economically feasible. Generally, and unlike 'Manufacture Defect', 'Design Defect' is not governed by strict liability); GOLDBERG & ZIPURSKY, *supra* note 47.

86. GOLDBERG & ZIPURSKY, *supra* note 47.

very different from an analysis of the reasonableness of the tortfeasor's decision: a design defect claim would require the victim to delve into the algorithm's overall programming in order to understand how it operated under different circumstances and why, including finding alternatives that could have been used in each of the phases of the algorithm's learning process, in order to show that a safer alternative existed.[87] This contrasts with the claim against the human physician, which would focus only on the specific damaging decision, and not on the general operation of the tortfeasor in various circumstances.[88] Secondly, and as will be elaborated on in Part IV, a design defect claim would probably have to address the programming aspects as well as the professional aspects of the decision-making process (i.e., what was included in the arsenal of programming tools that the programmer could have used; what was included in the arsenal of professional tools the algorithm could have chosen from).[89] By contrast, the claim against the human tortfeasor would focus only on the arsenal of professional tools he did or did not use.[90] Naturally, the extent of further costs and complexity, owing to the gulf between the underlying legal frameworks for the human and algorithmic tortfeasors, depends on the nature of the framework that would apply to algorithms (and as mentioned, even under a standard of "reasonableness" for both the human and algorithmic tortfeasors, different tests of "reasonableness" may exist and raise different questions). Nevertheless, in general, a unified framework applying to both cases would undoubtedly reduce the time and costs accruing in cases where joint tortfeasors are human and algorithmic.

### *3. Economic Distortion*

Human and algorithmic decision-makers are often supposed to fulfil a similar function, but each might have its particular advantages (and disadvantages). For example, in Amazon's "Prime Air" system, devised to make air deliveries by drones, a drone and a human deliverer alike might be put to use in moving shipments. But in certain circumstances, such as mountainous terrain, a drone might be more efficient, while in circumstances involving poor weather conditions, a human deliverer is better. A judge might have a relative advantage over a bail-setting algorithm; for instance, in cases of non-standard defendants or unusual circumstances where an algorithm might be misled because the information in the database it relied on was not relevant for the specific case in hand. On the other hand, a bail-setting algorithm might have the advantage of reaching numerous bail decisions in split seconds, thus saving the judicial system much time and money.

From an economic efficiency viewpoint, the different advantages of the two "production factors" render human and algorithmic decision-makers not entirely interchangeable. Instead, their optimal application would presumably

87. Hubbard, *supra* note 81, at 1821–22.
88. *Id.* at 1830.
89. *Id.* at 1821–22.
90. *Id.* at 1830.

require a mixture of both.[91] The precise optimal mixture depends on their cost-effectiveness, an equilibrium that is obviously affected by the costs associated with either factor. For example, a law firm wishing to expand might calculate that adding another attorney will yield annual returns of $300,000, while adding another Ross will yield annual returns of $350,000. The decision in favor of Ross is not automatic but depends on Ross's cost. If purchasing Ross is expected to cost $300,000 while a human attorney will cost $200,000, the latter will be more cost-effective (yielding a profit of $100,000 versus Ross's $50,000).

Everything else equal,[92] the application of a different tort framework to humans and to algorithms, which probably would result in a significant gap in the costs associated with litigation over damage each had caused, will affect the calculation and potentially result in fewer economic choices. For example, if the probability that a human attorney and Ross caused damage is equal, but the expected costs of litigation associated with damages caused by a human are $70,000 higher than Ross's, the firm will prefer Ross to a human.[93] This arbitrary gap in legal costs associated with either decision-maker will distort decisions and reduce efficiency. In other words, applying different legal frameworks will result in either algorithms or humans being used less often than they would have been based on their relative advantages and costs. Policy considerations may justify this outcome, but if so, these should be factored in deliberately.[94]

### *4. Chilling Effects*

Other than distorting economic decisions in specific cases, applying a different legal framework to human and algorithmic tortfeasors might negatively affect innovation and technological advancements in the field of autonomous algorithms.[95]

The playing field of algorithmic and human decision-makers is a-priori uneven in terms of chilling effects posed by litigation threats. Victims would probably be more likely to sue and judges more likely to "convict" algorithmic than human tortfeasors, in part because humans tend to sympathize with humans

91. *See, e.g., Review of Production and Cost Concepts*, MIT SLOAN SCH. OF MGMT. (Sept. 23, 2004), https://ocw.mit.edu/courses/sloan-school-of-management/15-010-economic-analysis-for-business-decisions-fall-2004/recitations/pro_and_cost_con.pdf.

92. See Abbot, *supra* note 26, at 112 (discussing the unequal parameters that contribute to the different costs of human labor versus machine labor).

93. The example assumes that, in both cases, the law firm is liable for the damages (and not, for example, the programmer or manufacturer of Ross if it caused the damage).

94. *See* Abbot, *supra* note 26, at 122 (focusing specifically on the differences between negligence that would apply to humans, and strict liability that Abbot assumes would apply to algorithms); *see also* Robert D. Cooter & Ariel Porat, *Lapses of Attention in Medical Malpractice and Road Accidents*, 15 THEORETICAL INQUIRIES L. 329, 352–53 (2014).

95. Naturally, the more easily-imposed and substantial the liability is, the more reluctant manufacturers or users would be to engage with it, thus leading to impeding development and innovation. *See, e.g.*, *Final Report Summary - ROBOLAW* (Regulating Emerging Robotic Technologies in Europe: Robotics Facing Law and Ethics), EUR. COMMISSION CORDIS, http://cordis.europa.eu/result/rcn/161246_en.html. This Article, however, does not analyze the general notion of "chilling effect" in the context of tortious algorithms, but rather focuses particularly on the potential chilling effect that might result from applying different tort frameworks on algorithmic and human wrongdoers.

and not with algorithms,[96] and because of other considerations such as prior acquaintance with the human (but not the algorithmic) tortfeasor or her relatives, fear of consequences or discomfort in future dealings with the tortfeasor's community, or a more forgiving attitude to human error. In addition, damages caused by autonomous machines are likely to receive much more media attention than similar damages caused by humans, thus causing the former more reputational damage and possibly leading to an increase of actions.[97]

Applying different tort frameworks to human and algorithmic decision-makers will, in certain cases, widen the gap in a way that might render the cost of autonomous algorithms prohibitive, and prevent or postpone technological progress.

First, assume that damages caused by a person or by an algorithm are to be borne by their employer—for example, a law firm in the case of Ross, or a hospital in the case of a robo-doctor. If the framework applicable to algorithms makes it easier and cheaper to collect compensation from algorithms than from human tortfeasors who have caused the same damage, the result is an increase in the relative cost of using algorithms as against humans.[98] This will influence an employer's decision to purchase an algorithmic decision-maker, and might also have a sweeping effect on the demand by employers for such algorithms, and therefore on the entire industry of algorithmic decision-makers in the given field. Such an outcome is to be expected in sectors where the demand for algorithms or for humans is set by the employer (who will bear the costs of damages) and not by end-users.

In sectors where customers will be free to choose which decision-maker to use—for example, in private clinics, where patients may choose to be treated by a human physician or by a robo-doctor—this analysis is reversed. Now, the easier it is to sue and collect from algorithms versus humans, the greater the end users' incentive to favor them and not their human counterparts. In other words, for these sectors, concern for diminishing demand for technology will arise when algorithms are more difficult to collect from, not the reverse.

Secondly, even when liability in the case of algorithms is to be borne by their manufacturers and not by their employers or users, lack of unity in applicable tort frameworks might incur additional costs for the 'manufacturers' of algorithms, thus creating a chilling effect, at least with respect to torts performed jointly by algorithms and humans, such as the cases discussed above. This is because under joint liability, victims are free to claim full compensation

96. Especially when the algorithms are not embedded in anthropomorphic technology such as 'human-like' robots (various studies found a connection between a human-like appearance of technology, and the level of engagement and compliance by human users). Byron Reeves & Clifford Nass, THE MEDIA EQUATION: HOW PEOPLE TREAT COMPUTERS, TELEVISION, AND NEW MEDIA LIKE REAL PEOPLE AND PLACES (1996); Clifford Nass Clifford & Scott Brave, WIRED FOR SPEECH: HOW VOICE ACTIVATES AND ADVANCES THE HUMAN-COMPUTER RELATIONSHIP (2005); Calo, *supra* note 31, at 55–58.

97. Karen Yeung, THE OXFORD HANDBOOK OF LAW, REGULATION AND TECHNOLOGY 538 (Oxford Univ. Press 2017); Ryan Calo, *Robotics & the Law: Liability for Personal Robots*, CIS (Nov. 25, 2009), http://cyberlaw.stanford.edu/blog/2009/11/robotics-law-liability-personal-robots.

98. Abbot, *supra* note 26, at 118.

from the wrongdoer of their choice.[99] If, for example, algorithms are easier to recover from, then victims will probably choose to sue the algorithmic decision-makers for the entire damage caused jointly by the algorithmic and human tortfeasors, even when the algorithm is responsible for only a small fraction of the damage. It is true that the proportional damage attributed to the human wrongdoer could be recovered from her in a separate action,[100] but such a step would mean additional litigation costs and risks that would accrue to the algorithmic decision-maker. The exact tipping point where developing and marketing algorithms become prohibitively expensive depends of course on myriad parameters. However, imposing additional costs on algorithm manufacturers owing to damages caused by human counterparts—which could be prevented or minimized were both algorithms and humans subject to the same framework—surely contributes to a chilling effect.

The chilling effect in the different fields of autonomous algorithms is of special concern. First, since we generally expect many of the developments in those fields to be incremental,[101] forestalling progress might have long-term effects that cannot be overcome soon after the causes of the chilling effect are removed. Secondly, delaying or preventing progress in those fields is troubling since algorithms are likely to outperform humans in various decision-making processes, thereby saving lives and resources.[102] Autonomous algorithms which can crunch enormous amounts of data and make unexpected cross-references from an almost unlimited number of sources, are expected to "unclog" the bottleneck of humans' limited abilities, and perhaps lead to many rapid game-changing discoveries. Since advancement in the field of algorithmic decision-making is exponential,[103] the human race likely has a great interest in allowing the development of such technologies to prosper.

Indeed, such innovation and progress also gives rise to the existential fears of placing too much power in the "hands" of machines, which may reach a point of no return where the 'Golem' would rise against its creator.[104] Moreover, without proper regulation, the wrong kind of innovation might be encouraged (e.g., algorithms that are cheap to operate but not accurate or reliable, in contrast to humans). These potential negative effects must therefore be carefully

99. *See, e.g.*, Richard W. Wright, *The Logic and Fairness of Joint and Several Liability, in Symposium, Comparative Negligence*, 23 MEM. ST. U. L. REV. 45, 45–46 (1992) (showing that under the joint and several liability doctrine, a plaintiff may pursue compensation from each liable party, as if the different parties were jointly liable for the tort. Plaintiffs wronged by multiple tortfeasors, in other words, may decide to file an action for the entire sum of compensation owed against any of the tortfeasors involved).

100. *Id.*

101. David C. Brock, *Reflections on Moore's Law*, UNDERSTANDING MOORE'S LAW: FOUR DECADES OF INNOVATION 87–104 (David C. Brock ed., 2006).

102. *See* Jeffrey K. Gurney, *Crashing into the Unknown: An Examination of Crash-Optimization Algorithms Through the Two Lanes of Ethics and Law*, 79 ALB. L. REV. 183 (2016) (containing a discussion of the many advantages of driverless vehicles).

103. *Id.*; Benjamin Alarie et al., *Law in the Future*, 66 UNIV. TORONTO L. REV. 423 (Nov. 7, 2016), https://ssrn.com/abstract=2787473.

104. Eur. Parl. Draft Rep. on Civ. L. Rules on Robotics, 2015/2103 (INL), at 4 (Jan. 1, 2017) ("[W]hereas ultimately there is a possibility that within the space of a few decades AI could surpass human intellectual capacity in a manner which, if not prepared for, could pose a challenge to humanity's capacity to control its own creation and, consequently, perhaps also to its capacity to be in charge of its own destiny and to ensure the survival of the species").

considered when developing the right mechanisms for creating better and safer technology, but so long as the importance and desirability of said technology is recognized, the chilling effect consequences of subjecting it to a different tort framework than the one that applies to humans needs to be taken into account.

### *B. Technology Neutral Standard*

Beyond resolving different anomalies resulting from the application of different legal frameworks to damages caused by algorithms or by humans, applying a reasonableness standard on algorithms is advantageous because it is neutral.

As mentioned above, algorithms' abilities are ever improving, to the point of mimicking and at times outperforming different human capabilities.[105] Undoubtedly, rapid and sweeping changes are expected in algorithmic decision-making, changes that will certainly affect the type, frequency, and magnitude of damages algorithms might cause.[106] In light of these changes, a "reasonable algorithm" standard is flexible and adaptable to the rate of development of the algorithm's abilities.[107] A normative standard of reasonableness may be adjusted to reflect different considerations that society wishes to promote with respect to algorithmic technology—for example, a desire for expedited innovation may lower the level of precautions needed for a decision to be reasonable,[108] and vice versa—while the general framework of reasonableness continues to apply without the need to reshape it in response to changes in technology and policy. A positive standard of reasonableness may too be adjusted, for example by comparing the actions of the algorithm to what other similar algorithms (and not persons) would do, once a sufficient number of similar algorithms are operational and thus provide comparative information.[109]

In addition to promoting statutory longevity and avoiding the need to constantly update laws and regulations,[110] a technology-neutral standard provides equal treatment of old and new technologies[111] and creates legal

105. Jason Millar & Ian Kerr, *Delegation, Relinquishment and Responsibility: The Prospect of Expert Robots*, *in* ROBOT LAW 102 (Ryan Calo et al. eds., 2016).

106. Jedrzj Niklas, *The Regulatory Future of Algorithms*, LONDON SCH. ECON. & POL. SCI. (Aug. 15, 2017), http://blogs.lse.ac.uk/ mediapolicyproject/2017/08/15/the-regulatory-future-of-algorithms/.

107. *See generally* Abbott, *supra* note 26 (discussing the reasonability standard with regard to algorithms and technology).

108. *Id*.

109. The damaging actions of "Ross," for example, may be analyzed in comparison to what a reasonable flesh and blood attorney would have done. But once some variety of similar lawyering algorithms exist in the market, Ross's actions may then be analyzed in comparison to what such algorithms would have chosen. Intuitively, the latter option seems like a more accurate reference point, but naturally a comparison to peer-algorithms would always be available only a while after the technology was first introduced.

110. John R. Kresse, *Privacy of Conversations over Cordless and Cellular Telephones: Federal Protection under the Electronic Communications Privacy Act of 1986*, 9 GEO. MASON L. REV. 335, 341 (1987) (discussing statutory longevity in this area).

111. Brad A. Greenberg, *Rethinking Technology Neutrality*, 100 MINN. L. REV. 1495, 1495–1562 (2016) (discussing the treatment of old versus new technologies, specifically advocating for "technological neutrality").

certainty.[112] A reasonableness standard applied to algorithms could therefore be advantageous in those contexts.

Overall, the adoption of human-like standards for algorithmic decision-makers entails certain advantages.[113] In addition to allowing flexibility and quick adaptation to technological advancements, it could have a positive effect on innovation and create proper incentives for the efficient use of humans and algorithms, and will allow victims of harm caused by algorithms to stand on equal ground in terms of recovery.[114]

## IV. ADDRESSING THE CONCEPTUAL DIFFICULTIES IN APPLYING A "REASONABLE ALGORITHM" STANDARD

Having reviewed the different anomalies and disadvantages, which may be the result of applying different tort frameworks on human and algorithmic tortfeasors, let us turn to examine whether the difficulties associated with applying a reasonableness standard to algorithms may be overcome at all. This Part reviews the different arguments against applying the "reasonableness standard" to autonomous algorithms, and addresses them.

A threshold argument against developing and applying reasonableness standards to algorithms is the mere fact that they are—well—not human.[115] Discussing tort liability of autonomous cars, Colonna, for example, discarded the notion of applying the negligence test to hardware or software. One of his main arguments was that "one of the integral pieces of the negligence analysis is deciding whether a "reasonable [person] of ordinary prudence" under like circumstances would have acted similarly. Yet . . . neither hardware nor software falls within the lay definition of a human being."[116]

As will be elaborated on below, the fact that algorithms are not human could and has been used in a variety of ways to argue that they do not warrant an independent analysis of reasonableness.[117] Firstly, the notion of examining the reasonableness of a machine might seem intuitively inappropriate or even peculiar.[118] Secondly, the fact that algorithms were programmed by humans begs the question of whether there is any meaning of analysing their own reasonableness separately from the reasonableness of their programmers (or, in other words, the question of whether the outcome of both "reasonableness"

112. Bert-Jaap Koops, *Should ICT Regulation Be Technology-Neutral?*, 9 STARTING POINTS FOR ICT REGULATION: DECONSTRUCTING PREVALENT POLICY ONE-LINERS, IT & LAW SERIES 77–108 (T.M.C. Asser Press 2006).

113. *See generally* Abbott, *supra* note 26 (discussing the reasonability standard with regard to algorithms and technology).

114. *Id.*

115. Saranya Vijayakumar, *Algorithmic Decision-Making*, HARV. POL. REV. (June 28, 2017), http://harvardpolitics.com/covers/algorithmic-decision-making-to-what-extent-should-computers-make-decisions-for-society/.

116. Colonna, *supra* note 26.

117. *See generally* Abbott, *supra* note 26 (discussing the reasonability standard with regard to algorithms and technology).

118. *Id.*

analyses might ever be different from one another).[119] Assuming that, indeed, there is no full equivalence between the reasonableness of the programmer and the reasonableness of the algorithm itself, another challenge to be dealt with is what the legal consequences of applying said analysis on algorithmic tortfeasors would be, given that algorithms lack legal status and cannot bear the consequences of their own actions.[120] In that respect, another difficulty is whether applying a reasonableness analysis to algorithms, therefore, could be in line with the rationales in the basis of tort law, which focus on deterrence and compensation of the victim.[121] Lastly, algorithms are, in many respects, superior to human decision-makers.[122] Even if we prove that there is a separate meaning to their own "reasonableness" and that said analysis could reconcile with tort law rationales, shouldn't we demand they meet an elevated level of performance than mere "reasonableness", which reflects human flaws but might be deemed too "forgiving" in the case of algorithms? This Part will address these challenges and concerns and show that reasonableness is, after all, a sensible mechanism to apply to algorithms.

### *A. Intuitive Reluctance*

The idea of subjecting algorithms to an analysis of their own "behaviour" or "choices" might seem improper. This is not only because machines are not expected (at least not in the near future) to be the ones legally responsible for the damages they had caused, but also because of an intuitive feeling that machines need not be "personified" and that legal tests or standards currently applied to humans should be reserved for them alone.[123]

Such perception is not necessarily warranted. First, it is not unprecedented that non-humans too are subject to an analysis of their own behaviour when determining tort liability (of other parties involved). [124] In more detail, canines are an example of non-humans (at least in the eyes of some), whose behaviour is analysed independently than the behaviour of their human owners in dog-attack cases, where liability is not imposed if the dog reacted 'proportionally' in response to a provoking act.[125]

119. Nanette Byrnes, *Why We Should Expect Algorithms to be Biased*, MIT TECH. REV. (June 24, 2016), https://www.technologyreview.com/s/601775/why-we-should-expect-algorithms-to-be-biased/ (discussing potential influence programmers have on their algorithms).

120. Tim Sprinkle, *Do Robots Deserve Legal Rights?*, AM. SOC'Y MECH. ENGINEERS (Jan. 2018), https://www.asme.org/engineering-topics/articles/robotics/do-robots-deserve-legal-rights.

121. Jules L. Coleman, RISKS AND WRONGS 197–212 (Cambridge Univ. Press, 1st ed. 1992).

122. Jason Collins, *What to Do When Algorithms Rule*, BEHAV. SCIENTIST (Feb. 6, 2018), http://behavioralscientist.org/what-to-do-when-algorithms-rule/.

123. *See* Sprinkle, *supra* note 120 (explaining that until we fully understand the implications of the technology, we should not afford it the same legal rights).

124. *See* Jay M. Zitter, *Intentional Provocation, Contributory or Comparative Negligence, or Assumption of Risk as Defense to Action for Injury by Dog*, 11 A.L.R. 5th 127 (2010) (analyzing one type of non-human, dogs, whose own behavior is considered when determining tort liability).

125. *Id.*

Moreover, and in general, intuitive perceptions on the desirable status of machines in the legal world may prove to be outdated.[126] Indeed, choices regarding legal recognition (such as granting legal status, legal standing, etc.) of non-humans have become flexible with time, and choices that were considered unthinkable at the time are beyond questioning nowadays.[127] Despite a previous perception that only persons could be eligible as a separate legal entity, for example, the law has acknowledged the legal status of companies, of municipalities and state bodies and even of vessels.[128] In other words, assumptions regarding the legal status or capacity of different entities which appeared to be undisputable, were often disproved completely along with different paradigm shifts.[129] Indeed, acknowledging the legal status of autonomous algorithms may be just around the corner, to judge based on the European Parliament recommendation discussed above.[130] Whether said proposal is ever accepted or not, the conceptual leap required for subjecting autonomous algorithms to negligence tests (or other legal mechanisms once reserved for humans) might very well seem an obvious one in hindsight.[131]

### *B. The "Homunculus Fallacy"*

As discussed in Part II, bot supervised and unsupervised machine learning is based on inputs given to the algorithm by its human programmers.[132] If algorithms make their choices according to how they were programmed to choose, what sense would there be in examining their own reasonableness, independently of their programmers? Balkin's "Homunculus Fallacy" argument refers precisely to this point: according to Balkin, "there is no little person inside the program."[133] Instead algorithms act as they are programmed to act—no more, no less.[134]

As elaborated below, programmers indeed affect their algorithm's choices, and may also prevent it from choosing certain alternatives altogether.[135] Still,

126. *See* Christopher D. Stone, *Should Trees Have Standing?—Toward Legal Rights for Natural Objects*, 45 S. CAL. L. REV. 450, 457, 471 (1972) (exploring the evolution of the legal status of different non-humans).

127. *Id.*

128. *Id.*; *see also* United States v. Brig Malek Adhel, 43 U.S. 210 (1844) (discussing a case where a vessel was involuntarily sold because it was used for illegal purposes by pirates who took over the ship. Vessel owners' claims that they never agreed to the illegal actions and should therefore not be punished were discarded, as the court explained that: "This is not a proceeding against the owner, it is a proceeding against the vessel for an offense committed by the vessel") (quoting United States v. Schooner Little Charles, 26 F.Cas. 979, 982 (Va. Cir. Ct. 1818)).

129. *See* Christopher D. Stone, *Response to Commentators*, 3 J. HUM. RTS. & ENV'T 100, 100–01 (2012) (explaining that in old civilizations, for example, imposing legal liability on slaves was considered foolish, given the clear differences between them and "real humans").

130. *See* European Parliament Draft Report, *supra* note 23 (suggesting that algorithms should be recognized under the law as separate from their human creators).

131. Stone, *supra* note 128.

132. Jason Brownlee, *Supervised and Unsupervised Machine Learning Algorithms*, MACHINE LEARNING MASTERY (March 16, 2016), https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/.

133. Balkin, *supra* note 28.

134. HERBERT A. SIMON, *The Corporation: Will It Be Managed by Machines?*, THE WORLD OF THE COMPUTER (J. Diebold ed., 1973).

135. Byrnes, *supra* note 119.

there is relevance and meaning in analysing an algorithm's own reasonableness separately from the reasonableness of the human programmers who chose to program the algorithm as they did, for the reasons presented below.

### *1. Unpredictable Outcomes*

A tortfeasor would breach the duty of care if she failed to adhere to the standard of reasonable care when carrying out actions that might foreseeably harm others.[136] While human developers certainly influence the choices of the algorithms they design, self-learning algorithms' choices are often wholly unforeseeable.[137] The makers of a coffee machine, for example, can foresee all the different scenarios the machine would possibly face and decide in advance the desired result for each one (heating the water when the user presses a certain button, pouring the water when it reaches a certain temperature, etc.). Programmers of self-learning algorithms, on the other hand, are in a completely different place in terms of foreseeing the results of their algorithms, which are "unpredictable by design."[138]

First, self-learning algorithms are frequently designed to outsmart the limits of the human mind, and draw conclusions that are beyond human comprehension.[139] Naturally, the more complex the algorithmic models are, the more difficult it is to understand and foresee such algorithms' choices.[140] "Deep-learning,"[141] for instance, poses a great challenge to programmers' ability to explain the "weights learned in a multilayer neural net" by the algorithm.[142]

---

136. RESTATEMENT (THIRD) OF TORTS: LIABILITY FOR PHYSICAL AND EMOTIONAL HARM § 3 (2010); Zipursky, *supra* note 54.

137. Millar & Kerr, *supra* note 105.

138. *Id.*

139. *See* Paulo J. G. Lisboa, *Machine Learning Approaches are Alone in the Spectrum in Their Lack of Interpretability*, *in* INTERPRETABILITY IN MACHINE LEARNING PRINCIPLES AND PRACTICE (Springer ed., 2013) (describing the interpretability of various machine learning approaches); *see also* Rodney Brooks, FLESH AND MACHINES: HOW ROBOTS WILL CHANGE US (Dan Frank & Stefan McGrath eds., 2003) (analyzing how robots and machine-learning can change humanity); Calo, *supra* note 31 (analyzing how robots can affect Cyberlaw).

140. *See generally* Jenna Burrell, *How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms*, 3 BIG DATA & SOC'Y 1 (2016) (explaining the difficulty in understanding algorithms used by complex machine learning algorithms); Will Knight, *The Dark Secret at the Heart of AI*, MIT TECH. REV. (Apr. 11, 2017), https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/ (explaining how as technology advances, understanding and communicating with intelligent machines becomes complicated).

141. Deep learning is a category of machine learning processes, based on unsupervised learning deriving information from myriad levels of information. Dong Yu, DEEP LEARNING: METHODS AND APPLICATIONS 200–01, (2014) https://www.microsoft.com/en-us/research/publication/deep-learning-methods-and-applications/.

142. "[I]t stands to reason that an algorithm can only be explained if the trained model can be articulated and understood by a human… There is of course a tradeoff between the representational capacity of a model and its interpretability, ranging from linear models (which can only represent simple relationships but are easy to interpret) to nonparametric methods like support vector machines and Gaussian processes (which can represent a rich class of functions but are hard to interpret). Ensemble methods like random forests pose a particular challenge, as predictions result from an aggregation or averaging procedure. Neural networks, especially with the rise of deep learning, pose perhaps the biggest challenge—what hope is there of explaining the weights learned in a multilayer neural net with a complex architecture?" Bryce Goodman & Seth Flaxman, *European Union Regulations on Algorithmic Decision-Making and a "Right to Explanation*" (Aug. 31, 2016) (presented at 2016 ICML Workshop on Human Interpretability in Machine Learning), https://arxiv.org/pdf/1606.08813.pdf.

Likewise, the tool of "randomness," which characterizes many self-learning algorithms, naturally renders their choices unforeseeable.[143]

Secondly, many machine-learning algorithms are online-based and may update their prediction models after each decision they make.[144] Unless the "person in the loop" needs to authorize each and every decision by the algorithm, the algorithm would reach conclusions based on new information that the human programmer would not have had a chance to consider.[145]

Granted, the programmer can certainly set boundaries that the algorithm may never cross (for instance, a robo-lawyer may be instructed never to make use of deceitful information), or instruct the algorithm to discard certain parameters altogether.[146] The human programmer may also maintain control over the algorithm by pre-programming it such that any significant change in its decision-making process will be subject to the programmer's approval.[147] For the algorithms to be useful, however, they cannot simply "freeze" whenever they encounter new information and change their conclusions accordingly, until the human programmer has had the chance to review and approve their suggested course of action.[148] A driverless car, for example, can simply not function if it must halt and wait for input from the programmer each time it encounters unfamiliar terrain.[149]

Thirdly, algorithms making complex decisions that replace human discretion (even on a non-self-learning basis) are expected to be unexpected. One reason is that complex decision-making flows are often programmed by a very large group of programmers, each contributing a certain amount of code lines and none having the ability of "seeing the whole picture" and able to predict the system's choices in each given scenario.[150] In addition, such algorithms often base their decisions on an astronomical number of combinations, each in turn containing an astronomical number of parameters and of potential

143. Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. PA. L. REV. 633, 653–56 (2017).

144. *Id.* at 660.

145. *Id.* ("[O]nline machine learning systems can update their model for predictions after each decision, incorporating each new observation as part of their training data. Even knowing the source code and data for such systems is not enough to replicate or predict their behavior—we also must know precisely how and when they interacted or will interact with their environment.").

146. Algorithms in the U.S. criminal justice system, for example, do not take into account variables such as race or gender. Such variables may nevertheless affect the algorithm's decision because other variables that are included in the model serve as proxies for race or gender. Angéle Christin et al., *Courts and Predictive Algorithms*, DATA & CIVIL RIGHTS, (Oct. 27, 2015) http://www.law.nyu.edu/sites/default/files/upload_documents/Angele%20Christin.pdf. A programmer may, however, carefully choose the database they "train" their algorithms on, to cause the algorithms to be less affected by biases. Anupam Chander, *The Racist Algorithm?* 115 MICH. L. REV. 1023, 1044 (2017).

147. Anupam, *supra* note 146, at 1045.

148. Kroll et al., *supra* note 143, at 699–700.

149. *See generally id.* (describing how computer scientists view these types of "oracles" in algorithms).

150. The Facebook Feed algorithm, for instance, which is surely less complicated than that of a robo-doctor, consists of numerous code lines programmed by various programs. Whenever the company decides on a change in the algorithm feed and some code lines are altered accordingly, the team, admittedly, do not know how the change will affect the algorithm and its choices. Rather, to implement a change, many trial and errors experiments are required. Will Oremus, *Who Controls Your Facebook Feed*, SLATE (Jan. 3, 2016, 8:02 PM), http://www.slate.com/articles/technology/cover_story/2016/01/how_facebook_s_news_feed_algorithm_works.html.

mixtures—far above what could be tested.[151] This renders it impossible for a programmer (even assuming there was only one programmer) to predict the choices made by the system under all potential scenarios.[152] Rather, only a very limited subset of scenarios could be tested to predict the choices they would yield.[153] Also, as mentioned above, many parameters could be dynamic and ever-changing.[154] Adding to the unpredictable nature of complex decision-making algorithms is their interaction with other inputs[155] and with other unpredictable codes as part of the Internet of Things revolution, allowing machines to communicate with themselves directly, without human involvement.[156] This makes it practically impossible to predict the algorithm's choices.[157]

In sum, while the choices of the older generation algorithms mirrored the choices made by their human programmers, autonomous algorithms now make choices that are not programmed by their developers and that are not foreseeable by them, at least with respect to a large subset of all possible scenarios. Therefore, there is no perfect equivalence between programmers' and algorithms' choices, and the algorithms' choices have an independent meaning.

---

151. Kroll et al., *supra* note 143 (explaining why both statistical and dynamic methods for reviewing algorithms in order to detect unwanted outcomes can only be effective for a small subset of all possible scenarios. This can be demonstrated by examples of different codes reviewed by many professionals, yet still containing unnoticed errors and "bugs," or by the famous proof by Alan Turing that "[t]here is no single algorithm that can predict whether, for any given program and input, the program will finish running at some point (halt) or will run forever." Alan Turing, *On Computable Numbers, with an Application to the Entscheidungsproblem*, 42 PROC. LONDON MATHEMATICAL SOC'Y 230 (1937)). Naturally, there are solutions for improving the chances of detecting errors in a code (such as dividing it to separate modules or annotating it with the assertions made while writing it). But this does not change the unpredictable nature of most of the choices made by sophisticated decision-making algorithms. *Id.*

152. *See* Kroll et al., *supra* note 143, at 650 n. 49 (describing how programs cannot be tested for every scenario).

153. Google, for example, admits that their driverless cars could not be pre-programmed for each possible scenario, given the astronomical amount of parameters to consider. Ari Shapiro, *All Things Considered: What Do Self-Driving Cars Mean for Auto Liability Insurance?*, NPR (Mar. 1, 2016, 4:20 PM) https://www.npr.org/2016/03/01/468751708/what-do-self-driving-cars-mean-for-auto-liability-insurance. The technology must be, therefore, based on observations and generalizations done by the machine itself which, in turn, leads us to the unpredictability inherent to machine-learning capabilities.

154. Such as, in the case of a robo-doctor, the exact body temperature of the patient, the level of pollution on the day of diagnosis, whether a certain epidemic is spreading in different parts of the world, etc.

155. If, to continue with our robo-doctor example, the robo-doctor downloads different software or programs, or is continuously updated with new medical discoveries and studies.

156. Nicholas Gane et al., *Ubiquitous Surveillance: Interview with Katherine Hayles*, 24 THEORY, CULTURE & SOC'Y 349, 350 (2007) (stating that "most of the communication will be automated between intelligent devices. Humans will intervene only in a tiny fraction of that flow of communication. Most of it will go on unsensed and really unknown by humans."); Katherine Hayles, *Unfinished Work: From Cyborg to the Cognisphere*, 23 THEORY, CULTURE & SOC'Y 159, 161 (2006) (stating that "[i]n highly developed and networked societies . . . human awareness comprises the tip of a huge pyramid of data flows, most of which occur between machines.").

157. A recent troubling example of just how surprising and unpredictable algorithms may be is demonstrated by the chatbots created by Facebook's A.I. research lab, who were "caught" conversing with each other in their own secret language. Chris Perez, *Creepy Facebook Bots Talked to Each Other in a Secret Language*, N.Y. POST (Aug. 1, 2017, 12:45 AM), http://nypost.com/2017/08/01/creepy-facebook-bots-talked-to-each-other-in-a-secret-language/; *see also* Calo, *supra* note 31, (suggesting that the notorious 2010 market "flash crash" was the result of the interaction between a few programs that responded unpredictably to each other).

*2. Time Lapse*

As mentioned, a tortfeasor breaches her duty of care if she does not adhere to the standard of reasonable care when carrying out actions that might foreseeably harm others.[158] Foreseeability, of course, is examined at the time the relevant action takes place.[159] Since certain harms may be unforeseeable when the algorithm is programmed, but foreseeable when the algorithm has eventually made its decisions, the human developer's choices and the autonomous algorithm's choices are once again not fully equal.[160]

In more detail, a coffee machine designed to heat and pour coffee upon receiving a certain input could continue its exact circle of operation for decades (perhaps a more efficient method of heating or pouring coffee will be discovered and will render the machine useless, but generally we do not expect any change in the foreseeability of different harms that the coffee machine might cause).

On the other hand, the time lapse in the case of an autonomous algorithm might in certain cases have significant consequences for whether certain harms are foreseeable. For instance, the programmer of a robo-doctor who in the past decided not to program "smallpox" as a medical condition would probably be deemed to have acted reasonably in doing so, because this disease was eradicated decades ago. If, however, the robo-doctor failed to diagnose smallpox a year later, after a smallpox epidemic spread across the country, then under every reasonableness criterion it would be deemed unreasonable.[161]

Granted, we could always go back to the programmer and check whether it was reasonable or not to have the programmer update the robo-doctor or have the robo-doctor refresh its knowledge on its own,[162] or whether the programmer should have issued a post-sale warning regarding the algorithm's discrepancies.[163] But, that analysis would focus on the programmer's reasonableness at the time the algorithm was programmed, based on the technology and sources of information available then. The analysis and its outcome may be different when focusing on the reasonableness of the algorithm

158. Miller & Perry, *supra* note 17, at 325.

159. RESTATEMENT (THIRD) OF TORTS: LIAB. FOR PHYSICAL HARM § 29 cmt. j (Proposed Final Draft No. 1, 2005).

160. Kroll et al., *supra* note 143, at 680.

161. This is because under a positive standard of reasonableness, presumably every physician in the country would have considered said diagnosis. Under a normative standard of reasonableness, too (to choose an economic one), the costs of considering the diagnosis of smallpox would likely be lower than the damage expectancy of misdiagnosing it.

162. In the context of driverless cars, for example, "over the air" updates of the vehicle's software were already successfully performed, including substantial updates relating to the car's movement. *See, e.g.*, Mark Rechtin, *Tesla Nimbly Updates Model S Over the Air*, AUTOMOTIVE NEWS (Jan. 16, 2013), http://www.autonews.com/article/20130116/OEM06/130119843/tesla-nimbly-updatesmodel-s-over-the-air (discussing substantial "over the air" updates); Damon Lavrinc, *In Automotive First, Tesla Pushes Over-the-Air Software Patch*, WIRED (Sept. 24, 2012), http://www.wired.com/autopia/2012/09/tesla-over-the-air/ (discussing "over the air" software updates).

163. As is required under product liability laws, assuming the "seller" knew or should have known of the substantial risk posed by its product. RESTATEMENT (THIRD) OF TORTS § 10(b) (AM. LAW INST. 1998). *See also* Bryant W. Smith, *Proximity-Driven Liability*, 102 GEO. L.J. 1777, 1808 (2014) (suggesting that "[u]ltimately a lack of remote updatability may itself constitute a design defect").

itself at the later time, when it made its decisions.[164] This once again supports the conclusion that a separate reasonableness analysis for the algorithm itself has a value distinct from that of the programmer.

### *3. Different Standards of Reasonableness*

The reasonableness of an algorithm's choices versus the programmer's choices would likely need to be evaluated under different standards. While the programmer is to be judged according to a "reasonable programmer" standard, the algorithm is to be judged by a standard of reasonableness pertaining to the specific field in which the algorithm is relevant.[165] It would make sense to assume, for example, that the standard of the "reasonable programmer" who programmed a robo-doctor or a driverless car would not be identical with the standards of the "reasonable robo-doctor" or the "reasonable driverless car." A robo-doctor, for instance, would presumably be judged according to the medical choice it reached as against the other medical alternatives.

On the other hand, the appropriate standard of care for the programmer is that of a "reasonable programmer."[166] Granted, if the programmer is designing professional robo-doctors, we would expect its programming skills and choices to be in line with those required to program a functioning robo-doctor, including, of course, making the right medical choices. We would, however, compare its choices with those of other programmers or with the opinions of experts in the field of programming, not the field of medicine. On a normative level, we would make cost-benefit analyses focused on the availability and efficiency of other programming means, rather than medical means.

To make this argument more concrete and to give specific examples of when an algorithm's choices might be deemed reasonable, whereas its programmer's decision to so choose would be deemed unreasonable, and vice versa, let us look at the following examples:

A battered woman and her spouse enter the clinic. The woman tells the physician she hit the door. While her injury matches that description, the physician notices something suspicious about her body language, and the way she avoids eye contact with her spouse. The spouse's reactions are also less (or more) sympathetic than expected, and the physician decides to follow the protocol that states that suspicion of domestic violence should be resolved by inviting a social worker to question the patient.[167] Assume also that to the human eye the suspicion described above is obvious, and had the couple visited a hundred physicians, all of them would have called a social worker. A robo-doctor, on the other hand, would have tremendous computational abilities and

164. Depending, of course, on the specific content poured into the reasonableness standard.

165. Patrick E. Hubbard, *'Sophisticated Robots': Balancing Liability, Regulation and Innovation*, 66 FLA. L. REV. 1803, 1855 (2014).

166. SIMON PRIEST ET. AL, EFFECTIVE LEADERSHIP IN ADVENTURE PROGRAMMING 153 (3d ed. 2017).

167. *See* City & Cty. of S.F. Dep't of Pub. Health, Community Public Health Services Domestic Protocol, http://www.leapsf.org/pdf/sample_clinic_protocol.pdf (providing protocol for domestic violence abuse).

could surely be programmed to detect suspicious vocal trembling;[168] but it could not necessarily replace human intuition as to when something is not as it appears. So, in this situation, a robo-doctor might not detect that something is wrong and would not request a social worker's intervention as required, perhaps thereby causing the woman to sustain another violent attack with significant physical injury. Is the robo-doctor acting reasonably when it fails to recognize the domestic violence situation? This depends, of course, on the standards of reasonableness we would apply. If, for example, we focus on the positive standard, and make the comparison with flesh and blood physicians, we already know that all the latter would act differently, therefore we can conclude that the robo-doctor was unreasonable. But, scrutinizing the programmer's reasonableness, and comparing her with other programmers, we might discover that there is no feasible method of designing a robo-doctor to detect accurately when 'something is wrong'; therefore, she acted as all programmers would act and was accordingly reasonable in doing so. Granted, applying a positive standard of reasonableness to a robo-doctor might require us to compare its actions with the actions of other robo-doctors, and not of human physicians, depending on the content of the standard of reasonableness to be developed for algorithms. In that case, if no robo-doctor in the world can detect when something is wrong, the equivalence between the robo-doctor's reasonableness and its programmer's is indeed sound. But, even if our positive comparison is made to a robo-doctor and not to a human physician, many robo-doctors in the industry may possibly possess the ability to detect when a patient's story is suspicious (for instance, if they are designed by an entity with inexhaustible human capital and monetary resources, such as a state or an army). In such a case our robo-doctor could still be found unreasonable under a positive standard of reasonableness. If, however, our programmer's more limited resources could, under no circumstances, allow her to develop such an ability of her robo-doctor, and if all similar programmers in the industry (not operating for a state or the army) have failed to add such a feature, we might view her programming as reasonable nevertheless. That is, even if we judge the reasonableness of non-human decision makers based on other non-humans' actions, we might still find them unreasonable while finding their programmers reasonable.

As mentioned above, medical malpractice is indeed usually judged according to a positive standard.[169] But, the example may still be accurate in other contexts analysed on the basis of a normative standard. If, for example, the cost of adding a feature that improves the robo-doctor's ability to recognize situations where the robo-doctor is lied to—or alternatively the cost of having the robo-doctor consult with a human in every case in order to make sure it is not "missing" anything—exceeds the expected utility, then the programmer is reasonable in avoiding said costs under the economic efficiency normative

168. *See* Susan Miller, *When Everybody Lies: Voice Stress Analysis Tackles Lie Detection*, GCN (Mar. 18, 2014), https://gcn.com/articles/2014/03/18/voice-risk-analysis.aspx (discussing the capabilities of voice recognition technology).

169. Philip G. Peters, *The Quiet Demise of Deference to Custom: Malpractice Law at the Millennium*, 57 WASH. & LEE L. REV. 163, 165 (2000).

standard. Likewise, if the programmer warns the users that the robo-doctor is not successful in identifying situations of untruthfulness, her behaviour is likely to be found reasonable. But at the same time, if the robo-doctor on a specific case could have made further inquiries, whose costs were lower than the expected damage, the robo-doctor might be found unreasonable.

A contrary example, that has to do with the time lapse argument discussed above, is when the algorithm's choices are deemed reasonable while its programmer's choices are not. Assume, for example, that our robo-doctor, now a pediatrician, has to diagnose and choose a course of treatment for a new-born whose brain seems underdeveloped. Considering different alternatives, the robo-doctor omits the fact that the baby suffers from microcephaly that resulted from the Zika virus, because the present scenario arose well before the outbreak of the virus, when only two or three sporadic cases were reported, all in very remote geographical areas. Even if the robo-doctor's failure to consider Zika results in choosing the wrong treatment and harms the baby, the robo-doctor is still likely to be found to have acted reasonably. This is under the positive standard of reasonableness: presumably all other human physicians, as well as robo-doctors, would at the time have discarded the option of a Zika infection; or under the normative standard of reasonableness based on economic efficiency, the great rarity of the virus at that time would have rendered it entirely uneconomical to test babies with microcephaly for Zika.[170]

The robo-doctor's programmer, however, might be found to have acted unreasonably if programming standards required that robo-doctors be continuously updated on all medical-related reports worldwide and automatically integrate them into their decision-making process.[171]

Though a clear correlation surely exists between the reasonableness of an algorithm and the reasonableness of its programmer, the two are not identical, and may occasionally lead to different results. Thus, an independent analysis of the reasonableness of an algorithm is not mere semantics, but could potentially yield two opposite outcomes.

### *C. Legal Implications and Reconciliation with the Rationales Behind Tort Law*

Having showed why an independent analysis of the reasonableness of an algorithm might warrant a different result than a similar analysis conducted with respect to the reasonableness of its programmer, the obvious challenge to address next is the legal implications of said analysis. Would it even mean anything, in practice, if an algorithm was found "reasonable" or "unreasonable"?

---

170. *Id.*

171. If this was the case, however, we would have probably expected more robo-doctors to take the Zika possibility into account, and thus render our robo-doctor unreasonable after all (at least when applying a positive standard of reasonableness) when failing to do so. But if, for instance, our robo-doctor was programmed at a later stage than all others in the industry (being a new model or something of that sort), when embedding continuous-updates of medical reports suddenly became technologically possible, the robo-doctor would still be deemed reasonable (acting like all other robo-doctors) while its programmer would not (as the option of adding said feature was readily available to the programmer).

This Part will mention, as food for thought, a few possible implications of said analysis. It will then concentrate on the implication that bears most similarities to the equivalent implication in the context of human tortfeasors, and examine whether the resulting legal consequences reconcile with the rationales of deterrence and of compensating the victim.

First, a finding that an algorithm acted "reasonably" or "unreasonably" might affect the way we judge the reasonableness of a human decision maker who chose to rely on an output or recommendation given by the algorithm. If, for example, a medical algorithm produced a damaging recommendation and a flesh and blood physician relied on it, then a finding that the algorithm's choice was reasonable would likely allow the physician to successfully argue that her reliance on the algorithm was in itself reasonable.[172] Since the current Article focuses on the reasonableness of the machine and not of the persons involved, let us move on to the second potential implication—one that would be of relevance in a future envisioned by the European Parliament[173]—where algorithms are awarded their own legal status. Under such a scenario, where algorithms are able to pay for their own damages, a finding that an algorithm acted reasonably or unreasonably would potentially have the same legal implications as would a finding of reasonableness by a person; reasonable algorithms would be resolved of liability, while unreasonable algorithms would have to pay for the damages they have caused.[174]

Until such day arrives, a third set of implications of the reasonableness (or unreasonableness) of an algorithm might affect the determination of liability of the party who currently assumes liability for damages caused by an algorithm. Indeed, even if the reasonableness of the algorithm itself is put to the test, the manufacturers, developers, or users of the algorithm may continue to pay for damages caused by it.[175] The fact that algorithms are no longer a "product," but operate at their own discretion should not matter in that context, just as employers pay for damage caused by their employees' negligence,[176] or, to give an example of a non-human tortfeasor, just as dog owners pay for damages caused by the behaviour of their animal.[177]

Regardless of the specific party that would assume liability for the damages caused by an algorithm said party is related to, there are several ways the "reasonableness" analysis may come into play under this set of alternatives, depending on policy considerations.

If policymakers wish to increase liability, they may add an "unreasonable algorithm" argument as a separate cause of action that might invoke liability by

172. Granted, the reasonableness of the physician would be analyzed based on the moment when she chose to rely on the algorithm's recommendation—a moment that naturally precedes the finding that the algorithm was indeed reasonable. But, assuming the recommendation was later found reasonable, it should be easier for the physician to show that her reliance on it was a choice other physicians would have also made (as discussed above, in general the standard of reasonableness in the world of medicine is a positive one, comparing a physician's actions to those of other doctors under similar circumstances).

173. Eur. Parl. Draft Rep. on Civ. L. Rules on Robotics, 2015/2103 (INL), at 6 (Jan. 1, 2017).

174. Abbot, *supra* note 26, at 104.

175. *Id.*

176. RESTATEMENT (SECOND) OF TORTS § 429 (AM. LAW. INST. 1965).

177. Duffy & Hopkins, *supra* note 1, at 469.

the humans involved. In other words, a victim harmed by a decision of an algorithm could either win her case on traditional causes of action (such as product liability or direct negligence by the humans or legal persons involved) or may prevail solely based on a finding that the algorithm was unreasonable.

If, however, policymakers wish to reduce the liability exposure of the relevant actors (for instance because of concerns of a chilling effect), then the unreasonableness of the algorithm may be a prerequisite for invoking other causes of action such as product liability or direct negligence. In other words, in order to recover damages, the victim not only would have to establish a "traditional" cause of action such as product liability, but would first have to show that the algorithm acted unreasonably, or else her case would be dismissed.

Naturally, both of these alternatives would continue to subject damages caused by algorithms to a very different legal framework than the one that currently applies to human tortfeasors: while the liability for human-caused damages is determined based on the reasonableness of the tortfeasor alone, the liability for algorithmic damages could be established by either reasonability of the tortfeasor or by other causes of action (according to the former alternative) or require that both are met at the same time (according the latter alternative). In that sense, both alternatives would not go to minimize the anomalies identified in Part III, and might in fact exacerbate them—adding a supplementary cause of action for victims damaged by algorithmic tortfeasors would place them at a better place compared to victims of human tortfeasors. It would also contribute to procedural inefficiency, as the underlying legal proceedings would then be comprised of both a traditional cause of action as well as an "unreasonableness" cause of action (which, as discussed above, is expected to raise different legal questions and entail additional costs). Naturally, paving the way for more successful proceedings against algorithmic damages by adding another cause of action would also contribute to the chilling effect of developing such algorithms in the first place, and would also increase economic distortion caused by under-usage of algorithms due to their expected high litigation costs.[178]

While the latter option of requiring both unreasonableness and an additional cause of action to be established in order to prevail in a lawsuit of algorithmic damages would likely not contribute to the chilling effect (as winning lawsuits against manufacturers would become more difficult), it would still place victims of algorithmic damage in an unequal place compared to other victims (who would only have to establish unreasonableness of the tortfeasor). It would also increase procedural inefficiency (under this alternative, all cases would have to address the question of both reasonableness and an additional cause of action) and in general might promote under-usage of human-decision

178. As discussed in Part II, the result might be the opposite when focusing on fields where the users themselves may choose between algorithmic and human decision-makers. In such cases, users would have an incentive to overuse algorithms, because redress in cases of damages would be easier for them.

makers (assuming proceedings of algorithmic damages would become more difficult to win).[179]

Though these alternatives certainly warrant more consideration, let us focus on an intermediate alternative, which bears most similarities to the current legal framework applying to human tortfeasors. Under this alternative, the analysis of reasonableness would be the only analysis used in order to determine liability (of the humans or legal persons involved). In other words, if the algorithm was deemed reasonable, no liability would be found. If, however, the algorithm was unreasonable, then the party sued, for example, its manufacturer, then the manufacturer would indeed be liable for the damages. Such a scenario would be most similar to the current legal status applying to human tortfeasors—as depicted in the table below.

| **OUTCOME FOR PLAINTIFF** | **TORTFEASOR IS REASONABLE** | **TORTFEASOR IS UNREASONABLE** |
|---|---|---|
| **HUMAN TORTFEASOR** | Plaintiff loses | Plaintiff wins |
| **ALGORITHMIC TORTFEASOR-ALTERNATIVE I** | Plaintiff may pursue additional causes of action | Plaintiff wins |
| **ALGORITHMIC TORTFEASOR-ALTERNATIVE II** | Plaintiff loses | Plaintiff might win if she establishes an additional causes of action |
| **ALGORITHMIC TORTFEASOR-ALTERNATIVE III** | Plaintiff loses | Plaintiff wins |

Applying a reasonableness standard on algorithms such that said analysis would on its own determine liability raises an additional challenge that stems from the fact that algorithms are not humans: would such an implication reconcile with the rationales behind tort law? The answer to said questions depend, like many things, on the details of the mechanism adopted, but in general, the answer might very well be a positive one.

### *1. Compensation*

Among the specific problems identified by Colonna in applying a negligence standard to software and hardware, is that non-humans cannot compensate for the damage they have caused, thus leaving victims uncompensated—in contrast with the rationales of tort law.[180] Even regardless of the European Parliament's suggestion to allow machines to pay for damage they had caused,[181] algorithms in fact do not need to pay themselves for the

179. *Id.* (stating that over-usage of human decision-makers in fields where users are free to choose between both types of decision makers).

180. Colonna, *supra* note 26, at 103.

181. European Parliament Draft Report on Civil Law Rules on Robotics, *supra* note 23, at § 31 (describing a shared obligatory insurance scheme or an individual funds for each robot category).

rationale of "compensation" to be met. As discussed above, the manufacturers, developers, or users of the algorithm may continue to pay for damages caused by their algorithms.[182] Focusing on the third alternative mentioned, then, while a finding that an algorithm was reasonable would resolve the "deep pockets" from liability—just as in the case of a human tortfeasor[183]—a finding that the algorithm was unreasonable (and met the other requirements of negligence)[184] would certainly not prevent victims from being compensated by the "deep pockets" involved. In fact, algorithms, especially those used in commercial contexts, are almost *always* expected to be linked to "deep pockets" in the form of their manufacturers or developers.

### *2. Deterrence*

Even with access to deep pockets for recovery, the fact that negligence is judged on the actions of the algorithms themselves might weaken the deterrent effect of tort liability, as the entity that caused the damage is not the same entity that has to bear the consequences, and hence would not internalize them when deciding how to act.

But even if algorithms themselves were not affected by a finding that they acted unreasonably, this still does not mean that they cannot be deterred from tortious conduct. True, a scenario in which algorithms themselves have "strong self-awareness" that causes them to "fear" the potential negative consequences of their actions (such as paying damages, if, in this futuristic world, algorithms owned assets) and shape their choices accordingly, still belongs to the world of science fiction.[185] Nevertheless, an algorithm may be programmed to consider the potential consequences of negligence as part of the parameters it weighs before reaching its decision.[186] For instance, if a robo-doctor needed to choose a certain treatment for a patient, the professional and economic parameters it would be programmed to weigh would include the damage expectancy associated with every possible alternative versus the costs of preventing damage.[187] So long as the entity that bore the costs of damages caused by an unreasonable algorithm could influence its decision-making process, the

182. Abbot, *supra* note 26, at 137.

183. RESTATEMENT (SECOND) OF TORTS § 429 (AM. LAW INST. 1965); Levin, *supra* note 19, at 1326.

184. An interesting question for a separate article might be whether and under which circumstances an algorithm owes a duty of care to potential users or third parties.

185. *See, e.g.*, A.I. ARTIFICIAL INTELLIGENCE (DreamWorks 2001); BICENTENNIAL MAN (Touchstone Pictures 1999) (providing some captivating examples of robots discovering that they experience human feelings).

186. Andrea Roth, *Machine Testimony*, 126 YALE. L. REV. 1972, 1991 (2017) (explaining how an algorithm may also teach itself to do so, if it discovers that such strategy advances it words a goal programmed into it).

187. *See, e.g.*, Brian K. Chen, *Defensive Medicine Under Enterprise Insurance: Do Physicians Practice Defensive Medicine, and Can Enterprise Insurance Mitigate its Effect*, (5th Ann. Conf. on Empirical Legal Stud. Paper, July 7, 2010) (describing how a robo-doctor would have considered the damage expectancy of each alternative as part of its professional decision making process, regardless of the threat of a tort law suit. But, aware of the fact they might be liable in torts for the damage resulting from an algorithm whose choices were unreasonable, the programmers of the algorithm may compute it to put emphasis on the legal consequences of a damaging choice, just like human doctors, when weighing their choices, often make their choices based on fear from malpractice suits).

rationale of deterrence would still apply in the same sense that deterrence works in the employer-employee relationship, for example, where employers are encouraged to direct their employees' behaviour such that it would not amount to a tortious act.

### *D. Would Superman be Subject to a Reasonableness Standard?*

Algorithms are superior to humans in many aspects of the decision-making process.[188] Algorithms, for instance, can compute an enormous amount of data that the human mind cannot grasp, let alone process within split seconds.[189] Unlike humans, algorithms do not have self-interests affecting their judgement,[190] they do not omit any of the decision-making stages or base their decisions on heuristics or biases,[191] and they are not subject to human physical or emotional limitations such as exhaustion, stress or emotionality.[192] Lastly, algorithms that are put to commercial or mass use will probably be among "the best ones out there," as against a human decision maker who may be above

188. Eben Harrell, *Managers Shouldn't Fear Algorithm-Based Decision Making*, HARV. BUS. REV. (Sept. 7, 2016), https://hbr.org/2016/09/managers-shouldnt-fear-algorithm-based-decision-making ("Algorithms tend to be superior to humans.").

189. Itiel E. Dror, *The Paradox of Human Expertise: Why Experts Get It Wrong*, *in* THE PARADOXICAL BRAIN 181 (Narinder Kapur ed., Cambridge Univ. Press 2011) (explaining how not only are human experts' abilities inferior in that sense, but enhancements of their performance might come at the expense of other abilities and degrade other aspects of their functioning).

190. *See* Gurney, *supra* note 102 (explaining how a human physician conducting research on a certain type of cancer, for instance, might unconsciously reach one diagnosis over another because the former would make the patient a candidate for her research. Unless programmed to take such considerations into account, an algorithm would not. This, in fact, is the root of the famous ethical "tunnel problem" dilemma, where a driverless car has to "decide" whether it would hit a child standing in the middle of the road or swerve into the tunnel and kill its passengers. When such a decision is made by a human driver, it is naturally affected by the driver's own self-preservation interest and instincts. An algorithm, on the other hand, is expected to consider such interests only if programmed to consider them.).

191. An algorithm, for instance, would not base its decisions on the "availability heuristic." The availability heuristic refers to people's tendency to base their estimations of the likelihood of certain events on prior knowledge that is easily retrievable. The more dramatic, emotional or unusual an event is, people tend to remember it better, and inaccurately base their estimations on it. *See* Amos Tversky and Daniel Kahneman, *Judgement Under Uncertainty: Heuristics and Biases*, 185 SCIENCE 1124, 1128 (Sept. 27, 1974). A human lawyer asked to predict a judge's reactions to a certain argument, for example, may recall the judge reprimanding her for raising a similar argument and will therefore overestimate the likelihood of the judge reacting negatively to said argument. An algorithm, on the other hand, will be affected by the database it "trained" on (and therefore if the data was not representative it would yield poor results), but, it would systematically analyze all prior incidents and give them an equal weight, without relying on a particular, emotional event. *See, e.g.*, Cass R. Sunstein, *Moral Heuristics* (U. Chi. L. & Econ., Olin Working Paper No. 180, Mar. 17, 2003), http://ssrn.com/abstract=387941; Cass R. Sunstein, *Hazardous Heuristics* (U. Chi. L. & Econ., Olin Working Paper No. 165, Nov. 23, 2002), https://ssrn.com/abstract=344620; Russel Korobkin, *The Problems with Heuristics for Law* (UCLA School of L., L. & Econ. Res. Paper No. 4-1, Feb. 9, 2004), http://ssrn.com/abstract=496462. It should be noted, however, that even if algorithms do not base their decisions directly on heuristics or biases, their decisions might nevertheless be affected by "hidden" biases (as is argued, for example, in the context of algorithms in the service of the criminal justice system. *See* Matthias Leese, *The New Profiling: Algorithms, Black Boxes, and the Failure of Anti-Discriminatory Safeguards in the European Union*, 45 SECURITY DIALOGUE 494 (2014); Toon Calders & Sicco Verwer, *Three Naïve Bayes Approaches for Discrimination-Free Classification*, 21 DATA MINING & KNOWLEDGE DISCOVERY 277 (2010); Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671 (2016)).

192. Philip Cooper, *Why AI Drives Better Business Decision-Making*, SALESFORCE: BLOG (Nov. 3, 2017), https://www.salesforce.com/blog/2017/11/why-ai-drives-better-business-decision-making.html.

average but, at the same time be the physician or attorney (for example) who finished last in their class.[193]

It could be argued that the entire concept of "reasonableness" stems from the inherent flaws of the human decision-making process, and that it is only because of these flaws that we do not demand perfect decisions at all times, but turn to a "second-best" level of desired human behaviour—which is "reasonable." If, for example, DC Comics' Superman were not fictional, the argument goes, we would not be satisfied with a "reasonableness" standard to judge his actions. Rather, we would probably expect and demand that Superman acts flawlessly at all times—simply because he can.[194] One could argue that for the same reasons, algorithms should not be subject to standards that stem from human weaknesses that they do not share.

The first response to this argument is that just as professionals are held to a higher degree of reasonableness than laypersons, and just as professionals with an established expertise are judged by standards higher than those applied to other professionals,[195] the standard of reasonableness for algorithms might be higher than that for a reasonable "person" or a "reasonable professional." Shaping an optimal reasonableness standard for an algorithm would clearly require much thought and adjustments, but the fact that it possesses superior abilities does not in itself render the standard of reasonableness irrelevant.

Secondly, and linked to the former point, algorithms are also inferior to humans in several aspects of the decision-making process. For example, they may suffer from technical malfunctions or be vulnerable to cyberattacks.[196] They also lack the creativity and flexibility that in many cases is a key component in becoming an expert in a professional field.[197] Moreover, algorithms may not necessarily be able to adjust their decisions when they encounter new parameters that were not part of their training process, or when such adjustments are not in line with their programmed limitations—unlike human decision makers, who might be better suited to manage unexpected circumstances or input.[198] Another aspect of algorithms' rigidity is that unlike human decision-makers they do not discuss their decisions with colleagues and peers and thus lack the opportunity to identify errors or tweak the decision so as to match the majority opinion.[199] Lastly, although algorithms' abilities outperform humans' in many aspects, they face an inherent disadvantage when

193. Khosla, *supra* note 9.

194. In terms of economic efficiency and optimal deterrence, if we go back to some of the rationales of tort law, it would not make sense to hold superman to standards of behavior that are clearly below his capabilities.

195. RESTATEMENT (SECOND) OF TORTS § 299A cmt. d. (AM. LAW INST. 1977).

196. Importantly, the negative outcome of an algorithm making erroneous decisions (due to malfunction of some sort or in general) could far exceed those of a human. This is because an algorithmic error might be duplicated to all other algorithms making that same decision, unlike a human whose decisions could vary from the ones made by other decision-makers. Lieblich & Benvenisti, *supra* note 45, at 30.

197. "Automation" is, in fact, a source of concern when assessing the performances of human experts, given that flexibility and creativity are essential for their functioning. Dror, *supra* note 189, at 182.

198. Lieblich & Benvenisti, *supra* note 45, at 26–32; Thomas J. Barth & Eddy F. Arnold, *Artificial Intelligence and Administrative Discretion*: *Implications for Public Administration*, 29 AM. REV. PUB. ADMIN. 332 (1999).

199. Lieblich & Benvenisti, *supra* note 45 at 26–32; Barth & Arnold, *supra* note 198.

making decisions calling for certain human traits that algorithms are not yet capable of copying. For instance, they might be less accurate than humans in detecting nuances or human body language or gestures indicating that a person is deliberately feeding them inaccurate information. In sum, algorithms do have certain disadvantages in decision-making as compared with humans, which supports the need for a specific reasonableness standard that considers their inherent strengths as well as their weaknesses.

Thirdly, the "lack of tolerance towards reasonableness" argument may be relevant for "right and wrong" decisions that depend on computational abilities, such as the optimal vectors for landing an airplane, for example, but it is less applicable to discretional decisions that have no "right or wrong" answers that could have been anticipated in advance. A robo-doctor making a damaging decision, for example, might have reached its decisions flawlessly in terms of weighing and calculating all the relevant parameters, but later it turned out that an alternative treatment would have worked better for the particular patient. Similarly, while the exact same decision could have won many physicians' support, others might have chosen an alternative.[200] "Reasonableness" questions suit exactly these types of cases, the accuracy of the relevant computations notwithstanding.

Lastly, and as was discussed in Part II, if algorithms are held to Superman-esque standards of strict liability, where responsibility for the damage is imposed regardless of culpability, this may create a significant chilling effect on the development of new algorithmic technologies.

## V. CONCLUSION

The more "human-like" algorithms become in terms of discretion and decision-making, the more sensible it is to stop treating them like "tools" and to apply legal concepts previously reserved for humans. In this Article, it has been argued that with respect to damages caused by such algorithms, a legal framework of the "reasonable algorithm" could be applied to algorithms just as the concept of the "reasonable person" or the "reasonable professional" applies to humans.

Having identified the anomalies and disadvantages resulting from applying different tort frameworks to algorithms and humans causing similar damages, this Article explained how conceptual difficulties associated with applying a reasonableness standard to algorithms are overcome. In more detail, the Article has demonstrated that analyzing an algorithm's reasonableness separately from the reasonableness of its programmer is of value, because it might yield opposite results (owing to the fact that algorithms' choices are unpredictable by design, and that the reasonableness of the algorithm versus the reasonableness of the programmer are measured at different points in time, focusing on different standards or reference points of reasonableness). It also discussed the possible

200. Karni Chagal-Feferkorn, *I Am an Algorithm, Not a Product: Why Product Liability is Not Well-Suited for Certain Types of Algorithms* (forthcoming) (discussing the characteristics separating certain algorithmic decision makers from "products" and thus warranting a different tort framework than product liability).

legal implications of applying a reasonableness test to algorithms, and explained why this would not be in contradiction with the rationales behind tort law. Showing that a "reasonable algorithm" standard is a concept that deserves thoughtful consideration; future articles might delve into the content that ought to be poured into the standard, as well as to the desired ways of applying it.