Osgoode Hall Law Journal

*Volume 58, Issue 2 (Spring 2021)* Article 1

7-15-2021

# A Theory of Vicarious Liability for Autonomous-Machine-Caused Harm

Pinchas Huberman
*Yale Law School*



## Citation Information

Huberman, Pinchas. "A Theory of Vicarious Liability for Autonomous-Machine-Caused Harm." *Osgoode Hall Law Journal* 58.2 (2021) : 233-284.
DOI: https://doi.org/10.60082/2817-5069.3678
https://digitalcommons.osgoode.yorku.ca/ohlj/vol58/iss2/1

# A Theory of Vicarious Liability for Autonomous-Machine-Caused Harm


## Abstract

The possibility of autonomous-machine-caused harm generates doctrinal and theoretical challenges for assigning tort liability. With emergent capabilities, autonomous machines disrupt the structure of interpersonal rights and duties in tort law, framed by conditions of foreseeability and proximate causation. Where algorithmic processes are unintelligible, self-modifying, and unpredictable, the concern goes, algorithmic harms will be untraceable to tortious human agency. As a result, their costs will simply lie where they fall—on faultless victims. This outcome would be unfair and objectionable: A failure of tort's mechanisms of corrective justice means faultless victims would disproportionately bear the accident costs of autonomous machines. This article suggests that the doctrinal form of vicarious liability is a promising strategy to ground tort liability for autonomous-machine-caused harm. Human or corporate deployers should be held liable for tortious harm caused by autonomous machines in the course of deployment. In this account, autonomous machines constitute a novel legal category as pure legal agents without legal personhood. In reconceiving vicarious liability—and the legal classification of autonomous machines—the article seeks to promote commonsensical liability outcomes for autonomous-machine-caused harm, consistent with tort's doctrinal and theoretical structure of corrective justice.

# A Theory of Vicarious Liability for Autonomous-Machine-Caused Harm

PINCHAS HUBERMAN*

The possibility of autonomous-machine-caused harm generates doctrinal and theoretical challenges for assigning tort liability. With emergent capabilities, autonomous machines disrupt the structure of interpersonal rights and duties in tort law, framed by conditions of foreseeability and proximate causation. Where algorithmic processes are unintelligible, self-modifying, and unpredictable, the concern goes, algorithmic harms will be untraceable to tortious human agency. As a result, their costs will simply lie where they fall—on faultless victims. This outcome would be unfair and objectionable: A failure of tort's mechanisms of corrective justice means faultless victims would disproportionately bear the accident costs of autonomous machines.

This article suggests that the doctrinal form of vicarious liability is a promising strategy to ground tort liability for autonomous-machine-caused harm. Human or corporate deployers should be held liable for tortious harm caused by autonomous machines in the course of deployment. In this account, autonomous machines constitute a novel legal category as pure legal agents without legal personhood. In reconceiving vicarious liability—and the legal classification of autonomous machines—the article seeks to promote commonsensical liability outcomes for autonomous-machine-caused harm, consistent with tort's doctrinal and theoretical structure of corrective justice.

* Pinchas Huberman (J.D., LL.M., University of Toronto, Faculty of Law), LL.M., Yale Law School), incoming J.S.D. Candidate at Yale Law School. I am grateful to Professors Peter Benson and Bruce Chapman for valuable feedback and comments on earlier versions of this article, which constituted part of my LL.M. thesis at the University of Toronto. I am also grateful to the reviewers and editors at the Osgoode Hall Law Journal for helpful comments and suggestions, and to Ben Ohavi and Yona Gal for many lively discussions about ideas in this article.

DEVELOPMENTS IN ROBOTICS AND ARTIFICIAL INTELLIGENCE are triggering deployment of autonomous machines, or autonomous agents (AAs), in new roles, projected to be significant parts of our social fabric in coming years (*e.g.*, driving cars, performing surgeries, caregiving).[1] Technological advances have enabled computer and algorithmic systems to learn from experience by analyzing large amounts of instructive data at extraordinary speed, and to interact with their external environments by using sensors and actuators to perform dynamic physical tasks.[2] As a result, we expect increased interaction between humans and autonomous machines, presenting novel risks of accidental harm to individuals and property.[3] There have already been, in this respect, harmful accidents involving self-driving cars and warehouse robots, resulting in legal

1. See *e.g.* Meera Senthilingam, "Would You Let a Robot Perform Your Surgery By Itself?" (12 May 2016), online: *CNN* <www.cnn.com/2016/05/12/health/robot-surgeon-bowel-operation/index.html>; Adam Goldenberg, "If an Autonomous Vehicle Has an Accident, Who is Legally Responsible?" (18 December 2018) online: *Maclean's* <www.macleans.ca/opinion/if-an-autonomous-vehicle-has-an-accident-who-is-legally-responsible>; Adriana Barton, "Cyclons, They Are Not. These Intelligent and Friendly Robots Are Designed to Help the Elderly Live a Better Life" (26 August 2018) online: *Globe and Mail* < www.theglobeandmail.com/life/health-and-fitness/article- cylons-they-are-not-these-intelligent-and-friendly-robots-are>; Jacqueline Howard, "Robot Pets Offer Real Comfort" (1 November 2017) online: *CNN* <www.cnn.com/2016/10/03/health/robot-pets-loneliness/index.html>.
2. Jerry Kaplan, *Humans Need Not Apply: A Guide to Wealth and Work in the Age of Artificial Intelligence* (Yale University Press, 2015) at 4-5.
3. See *e.g.* Ryan Calo, "Robotics and the Lessons of Cyberlaw" (2015) 103 Cal L Rev 513 at 534 [Calo, "Robotics and Lessons of Cyberlaw"].

suits for personal injury and wrongful death.[4] These developments raise critical questions and uncertainty about the scope of human accountability for harms of autonomous machines using emergent algorithms. Where algorithmic processes are unintelligible, self-modifying, and unpredictable, the concern is that designers and users will lose full operative control of their outputs and effects, thereby reducing accountability for resultant accidental harms.[5] A growing body of tort scholarship, in particular, reveals significant concern about liability gaps: scholars worry that suppliers, owners, and users of autonomous machines will too easily escape liability because of difficulties in tracing algorithmic harms to tortious human agency.[6] As a result, victims would disproportionately bear the accident costs, which is, arguably, an unfair and objectionable liability outcome.

The possibility of autonomous-machine-caused harm, then, generates significant doctrinal and theoretical challenges for assigning tort liability. Autonomous machines have features that disrupt the structure of interpersonal rights and duties in tort, which is framed by conditions of foreseeability and

4. See *e.g.* Daisuke Wakabayashi, "Self-Driving Uber Car Kills Pedestrian in Arizona, Where Robots Roam" (19 March 2018) online: *New York Times* <www.nytimes.com/2018/03/19/technology/uber-driverless-fatality.html>; Peter Holley, "After Crash, Injured Motorcyclist Accuses Robot-Driven Vehicle of 'Negligent Driving'" (25 January 2018) online: *Washington Post* <www.washingtonpost.com/news/innovations/wp/2018/01/25/after-crash-injured-motorcyclist-accuses-robot- driven-vehicle-of-negligent-driving/?utm_term=.ca85154c515e>; The Fernandez Firm, "What Happens When a Robot Causes Wrongful Death?" (30 March 2017) online: *Medium* <medium.com/@thefernandezfirm/what-happens-when-a-robot-causes-wrongful-death-3d1f4f7e9711>; Seth Baum & Trevor White, "When Robots Kill" (23 July 2015) online: *The Guardian* <www.theguardian.com/science/political-science/2015/jul/23/when-robots-kill>.

5. See *e.g.* Andreas Matthias, "The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata" (2004) 6 Ethics and Information Technology 175; Jack M Balkin, "The Three Laws of Robotics in the Age of Big Data" (2017) 78 Ohio St LJ 1217 at 1233-34; Frank Pasquale, "Toward a Fourth Law of Robotics: Preserving Attribution, Responsibility, and Explainability in an Algorithmic Society" (2017) 78 Ohio St LJ 1243 at 1247-55; Madeleine Clare Elish, "Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction" (2019) 5 Engaging Science, Technology, and Society 40; Alexander Campolo & Kate Crawford, "Enchanted Determinism: Power Without Responsibility in Artificial Intelligence" (2020) 6 Engaging Science, Technology, and Society 1.

6. See *e.g.* Calo, "Robotics and Lessons of Cyberlaw", *supra* note 3; David C Vladeck, "Machines Without Principals: Liability Rules and Artificial Intelligence" (2014) 89 Wash L Rev 117; Peter M Asaro, "The Liability Problem for Autonomous Artificial Agents" (Association for the Advancement of Artificial Intelligence 2016 Spring Symposium Series delivered at the Stanford University, 22 March 2016), (AAAI Press, 2016); Curtis EA Karnow, "The Application of Traditional Tort Theory to Embodied Machine Intelligence" in Ryan Calo, A Michael Froomkin & Ian Kerr, eds, *Robot Law* (Edward Elgar Publishing, 2016) 51 at 74; Jack Balkin, "The Path of Robotics Law" (2015) 6 Cal L Rev 45 at 51-55.

proximate causation.[7] As a general matter, AAs can sense phenomena, process what they sense, and act upon the world.[8] AAs are embodied and emergent:[9] As embodied entities, AAs take some corporeal form, are designed to act in the world, and can directly impact individuals and property.[10] As emergent entities, their behaviours are not entirely pre-programmed. AAs can update their algorithms through machine learning techniques to effectively adapt to new circumstances.[11] With emergent capabilities, AAs operate with some degree of functional independence; they are empowered to self-select methods to achieve programmed goals.[12] There are, moreover, two dimensions to AAs' independence. First, AAs' machine learning capabilities and engagement with novel circumstances negate full control and foreseeability on the part of human designers and users.[13] Second, through machine learning processes, AAs' controlling algorithms are designed by their learning algorithms.[14] In this latter affirmative sense, AAs modify their behaviours through internally caused processes—a kind of functional agency.[15] AAs are still, no doubt, deterministic systems: their outputs are defined by inputs received.[16] AAs' ultimate goals are also not their own, but programmed by human designers to advance human interests. Nevertheless, due to AAs' emergence, it is exceedingly difficult to fully trace connections between environmental inputs and changes in their

7. See Pinchas Huberman, "Tort Law, Corrective Justice and the Problem of Autonomous-Machine-Caused Harm" (2021) 34:1 Can JL & Jur 1 105 at 109-113 (for a lengthier discussion of AAs' salient features threatening to disrupt tort law).
8. See Calo, "Robotics and Lessons of Cyberlaw", *supra* note 3 at 529. This is neatly phrased as the "sense-think-act paradigm."
9. *Ibid* at 532.
10. *Ibid* at 534.
11. *Ibid* at 538-39. See also Shannon Vallor & George A Bekey, "Artificial Intelligence and the Ethics of Self-Learning Robots" in Patrick Lin, Keith Abney & Ryan Jenkins, eds, *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence* (Oxford University Press, 2017) 338 at 340; Harry Surden, "Machine Learning and Law" (2014) 89 Wash L Rev 87 at 89-95; Deven R Desai and Joshua A Kroll, "Trust but Verify: A Guide to Algorithms and the Law" (2017) 31 Harv JL & Tech 1 at 26-29; Matthias, *supra* note 5 at 179.
12. See Karnow, *supra* note 6 at 56-60.
13. Matthias, *supra* note 5 at 182.
14. See Desai & Kroll, *supra* note 11 at 28.
15. See Ugo Pagallo, *The Laws of Robots: Crimes, Contracts and Torts* (Dordrecht: Springer Science + Business Media, 2013) 38; Wolf Loh & Janina Loh, "Autonomy and Responsibility in Hybrid Systems" in Patrick Lin, Keith Abney & Ryan Jenkins, eds, *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence* (Oxford University Press, 2017) 39.
16. See Neil M Richards & William D Smart, "How Should the Law Think About Robots" in Ryan Calo, A Michael Froomkin & Ian Kerr, eds, Robot Law (Edward Elgar Publishing, 2016) 3 at 18.

algorithms and behaviours.[17] In this sense, AAs are "unpredictable by design."[18] With combined capacities of emergence and embodiment, AAs' characteristic unpredictability is coupled with the potential to cause personal injury and property damage. Even well-trained AAs can produce undesirable outputs in the real world, including accidental harm.[19] While such harm reflects misalignment between AAs' codes and programmers' goals, this misalignment may not be introduced by, or knowable to, programmers themselves.[20]

Consequently, AAs' harmful effects may be principally untraceable to tortious actions of designers, manufacturers, or users.[21] If so, under traditional tort doctrine, the cost of harm resulting from AAs' emergent behaviours will simply lie where it falls.[22] Under the law of negligence—tort law's paradigmatic doctrine in the context of accidents causing personal injury or property damage—tort liability follows only if the plaintiff's loss is caused by, and within the scope of, a defendant's tortious act involving foreseeable and unreasonable risk to the plaintiff.[23] If AAs' emergent processes, outputs, and harmful effects are typically unforeseeable, it will be difficult to identify tortious conduct by their designers, manufacturers, or users.[24] Moreover, if AAs' harmful effects stem from machine-learning algorithms, impacted by variable environmental inputs, it will be difficult to attribute any particular instance of AA-caused harm to responsible human conduct as its proximate cause.[25] The harm may be more precisely attributable to some environmental input, rather than a tortious aspect of the AA's design, construction, or operation, if any. Finally, AA-caused harm is not analogous to common law strict liability categories, such as keeping wild animals, using explosives, or transporting gasoline. Deploying AAs should not be viewed as exceedingly dangerous. In some instances, AAs are projected to be less risky than the human actors they replace. For instance, automated vehicles will likely be safer than conventional vehicles due to the absence of

17. Matthias, *supra* note 5 at 182.
18. Calo, "Robotics and Lessons of Cyberlaw", *supra* note 3 at 542.
19. See Vallor & Bekey, *supra* note 11 at 343.
20. *Ibid*.
21. See Vladeck, *supra* note 6 at 121-23.
22. See Calo, "Robotics and Lessons of Cyberlaw", *supra* note 3, 542; Asaro, *supra* note 6, 191; Karnow, *supra* note 6, 63-74; F Patrick Hubbard, "Sophisticated Robots: Balancing Liability, Regulation and Innovation" (2014) 66 Fla L Rev 1803 at 1851-52.
23. *Overseas Tankship (UK) Ltd v Morts Dock & Engineering Co Ltd (The Wagon Mound)*, [1961] AC 388.
24. See Karnow, *supra* note 6 at 74; Pagallo, *supra* note 15 at 117. See also Karnow, *supra* note 6 at 63-64, 72-73.
25. See Asaro, *supra* note 6 at 191-92.

human-driver-error.[26] If deploying AAs is expected to reduce risks associated with human activities, it should not be characterized as an abnormally dangerous activity.[27] It is dissimilar to keeping wild animals or using explosives, which are inherently tortious, involving characteristic risk that exceeds the levels normally assumed in ordinary patterns of interaction.[28]

This doctrinal outcome reflects tort law's theoretical structure of corrective justice, which does not straightforwardly apply to AA-mediated social interaction.[29] As a practice of corrective justice, tort law assumes a bilateral relation as its subject: its doctrines determine whether there is a legally relevant link between a particular defendant's tortious act and the plaintiff's loss, grounding the defendant's agent-specific duty to repair.[30] In the context of personal injury and property damage, tort liability responds to breaches of qualified duties of relational non-injury:[31] where wrongful action materializes in loss to an object of the plaintiff's right. In this view, tort law assumes a relation between two legal persons situated symmetrically with correlative rights and duties, delineated by its doctrinal terms of fair interaction. Tort law demarcates the scope of rightful action with an objective standard of care, setting reciprocal constraints on individuals' actions. Acts of risk imposition that exceed an objectively reasonable level are deemed tortious: The law evaluates whether individuals' external actions are wrongful, in this respect, without judging individuals' inner-character, blameworthiness, or subjective lack-of-concern for others.[32]

Tort liability operates, moreover, only upon the conjunction of wrongful action and causation. Absent wrongdoing, causation of harm is not a tort: If the act is consistent with norms of rightful action, the causation of harm does not

---

26. Bryant Walker Smith, "Automated Driving and Product Liability" (2017) Mich St L Rev at 15.
27. Ryan Abbott argues, in this respect, that where autonomous machines are safer than their human counterparts, human actors should be subject to the more exacting standard of a "reasonable robot," the purpose of which is to deter the human activity and incentivize adoption of the preferable automated alternative. See Ryan Abbott, *The Reasonable Robot: Artificial Intelligence and the Law* (Cambridge University Press, 2020) at 66-70.
28. For a lengthier doctrinal analysis of AA-caused harm under the laws of negligence and strict liability, see Huberman, *supra* note 7 at 122-31.
29. For a more comprehensive presentation of this argument, see *ibid* at 118-22.
30. See Ernest J Weinrib, *The Idea of Private Law* (Oxford University Press, 2012) at 168-70 [Weinrib, *Private Law*].
31. See Arthur Ripstein & Benjamin C Zipursky, "Corrective Justice in an Age of Mass Torts" in Gerald J Postema, ed, *Philosophy and the Law of Torts* (Cambridge University Press, 2001) 214 at 218-20.
32. See OW Holmes, *The Common Law* (Little, Brown, and Company, 1881) at 108.

infringe upon the plaintiff's right.[33] Likewise, absent causation, a wrongful act is not a tort as, in this instance, there is no interference with an object of a plaintiff's right at all.[34] The locus of tort liability is wrongful loss, not just any loss suffered by innocent victims, nor any unreasonable activity undertaken by tortious actors. Tort liability is, then, concerned with a version of outcome-responsibility: Liability results from the normative attribution of a plaintiff's loss to a wrongful action as its legally responsible cause. As a matter of corrective justice, tort law implicates responsible human agency, grounding agent-specific liabilities and duties to rectify harmful effects of wrongful action.[35]

AA-caused harm does not fit neatly within this basic structure.[36] Since AAs are not legal subjects bearing tort duties, harm resulting from AAs' emergent processes—*i.e.*, processes which are not traceable to tortious programming, instruction, or manipulation by human designers or users—does not stem from pertinent legal agency. AA-caused harm, in this sense, seems to be legally comparable to natural-events-causing harm, which leaves victims without an avenue for compensation in tort. This outcome reflects a traditional application of tort law, which evaluates whether particular instances of AA-caused harm are legally attributable to human tortious conduct—whether negligent use or design. However, this traditional analysis implicitly views AAs as ordinary products, and fails to adequately grasp the distinctive character of AA-mediated social relations. AAs are expected to perform tasks previously undertaken exclusively by humans—*e.g.*, driving cars, performing surgeries, caregiving—with some functional independence. AAs' social significance is not entirely analogous to ordinary products: AAs act on environmental inputs that are not traceable to particular human designers or users, and their outward effects are not necessarily attributable to human agency, as the effects of products typically are. But AA-caused harm is also not morally equivalent to natural-events-causing harm: Unlike natural disasters, AAs are purposely deployed by humans to advance their interests. AAs have special social and normative significance, as mediators of human interaction: They are deployed by individuals and corporations to independently perform complex tasks in interaction with other individuals and property. AAs' special utility and functional independence suggest that they are

33. Peter Benson, "Misfeasance as an Organizing Normative Idea in Private Law" (2010) 60 UTLJ 731 at 766-70.
34. See Arthur Ripstein, *Private Wrongs* (Harvard University Press, 2016) at 116.
35. Jules L Coleman, "The Practice of Corrective Justice" in David G Owen, ed, *Philosophical Foundations of Tort Law* (Clarendon Press, 1995) 53 at 66-69.
36. For a lengthier discussion of the argument in this paragraph and the next, see Huberman, *supra* note 7 at 121-22, 135-37.

better regarded as quasi-social-actors, extraordinary entities without self-evident legal classification.

Construing AAs as (functional) actors reveals a critical insight: There is no tort liability for AA-caused harm under negligence or strict liability because distributors or users are not the relevant actors that cause the harm—AAs were deployed to perform the tasks instead. These doctrines do not capture the relevance of (reasonable) deployment of another actor to perform a task as grounds for liability. A formalistic application of tort law, then, leads to a fundamental incongruity: Whereas personally performing tasks causing accidental harm could potentially produce tort liability, individuals can avoid liability altogether by deploying AAs to perform the tasks instead. And, in turn, the emergent and harmful effects of AAs fall outside the realm of tort law as they are not legal subjects. This liability outcome, however, does not capture the normative significance of distributing or using AAs, which entails an uncontrollable potential of harm, at least in isolated cases. Since deployers utilize AAs' emergent features to advance their purposes, they should assume some responsibility for AA-caused harm—in the agent-specific sense, as a matter of tort law—despite reduced foreseeability and proximate causation. This demands creative use of tort categories and potentially reconceiving AAs' legal classification to evaluate AAs' proper normative impact on the rights and duties of legal persons.

To this end, this article considers the doctrinal form of vicarious liability for AA-caused harm: whether deployers (users or distributors) of AAs may be liable for tortious acts committed by AAs in the course of deployment.[37] It argues that a vicarious liability approach offers a pragmatic doctrinal solution that rests on plausible theoretical foundations. In this account, AAs' outward behaviours would be evaluated as tortious or non-tortious. AAs commit torts where their outward behaviours fail to conform to standards of care expected of reasonable persons in the circumstances. Vicarious liability then attaches to defendant deployers for tortious acts of AAs that occur in the course of deployment, circumventing the need to trace AA-caused harm to tortious conduct by deployers themselves (*i.e.*, negligent use or legally defective design). The solution is a pragmatic one: It is a way to discover a basis of liability despite AAs' functional independence and characteristic unpredictability, which prevents direct attribution of AAs' emergent processes, outputs, and harmful effects to human agency. It is also conceptually plausible and normatively sound: It reflects deployers' agent-specific

37. The deployers, for the purposes of vicarious liability, may be users or distributors, as discussed in Part III(D), below.

responsibility for deploying functionally independent and instrumentally rational actors to perform tasks entailing the distinctive risks of AA-caused harm.

The move to vicarious liability for AA-caused harm is certainly controversial. It implies that AAs are legal actors capable of committing torts that can be imputed to human principals. At the core of the vicarious liability account is a legal stance toward AAs that is analogous to employees or legal agents: AAs act on behalf of human or corporate deployers, empowered to ground deployers' liability for their tortious harms occurring in the course of deployment. This approach generates fresh doctrinal and theoretical complications to be addressed in this article: What is the method to evaluate AAs' actions as tortious or non-tortious? What is the method to determine whether AAs' tortious actions are sufficiently connected to their deployment to impute their tortious acts to their deployers? How does the agency relation between AAs and their deployers arise? Can AAs be coherently compared to employees for purposes of tort liability, though they are not legal persons?

This article proceeds as follows. Part I reviews the doctrinal elements of common law vicarious liability: a tort, committed by an employee, in the course of employment. It focuses, specifically, on the doctrinal role of enterprise risk, which forms the requisite link between employees' torts and their employers, grounding vicarious liability. Part II provides a conceptual account of vicarious liability, consistent with its doctrinal configuration and tort law's basic structure, which embodies agent-specific reasons for liability. In this account, vicarious liability reflects employers' and employees' joint production of, and joint responsibility for, tortious harm. Where an employee's tortious act is linked to risks of enterprise, the act is also attributable to the employer: Since the employer enterprise is identified with the employee's tortious act, it is deemed an additional responsibility base. In this view, vicarious liability grounds agent-specific reasons for liability specifically suited to employment relations in which employers and employees act jointly in pursuit of collective aims of enterprise. This account lays the foundation for conceiving of AA deployment as a form of legal agency—akin to employment—as elaborated upon in the following section.

Part III extends the doctrinal elements of vicarious liability to the context of AA-caused harm. AAs are deemed to be legal actors pursuant to a legal intentional stance, with capacity to commit tortious acts. AAs' tortious acts are then attributable to their deployers when committed in the course of deployment—that is, when linked to characteristic deployment risk—which grounds deployers' vicarious liability. A distinctive theoretical conception of legal agency is at play in the vicarious liability account: AAs are pure legal agents without legal

personhood. AAs can therefore give effect to legal consequences for deployers, but not on their own account. This article argues that a pure legal agency classification nicely captures AAs' social and normative position within human relations, as functionally independent and rational instrumentalities deployed to act exclusively for deployers' purposes. The pure legal agency classification emphasizes the legal significance of the deployment relation as a whole—not AAs' independent legal or moral status—implicating deployers' responsibility for AAs' outputs and harmful effects.

This account, moreover, is illuminated by philosophical accounts that perceive sophisticated technologies as extending human agency. Drawing on the philosophy of technology scholarship of Deborah Johnson and F. Allan Hanson, it argues that AAs' tortious acts cannot be dissociated from their deployment and its characteristic risks. AAs' tortious acts committed in the course of deployment should stand in as surrogates to evaluate deployers' responsibility in tort. Crucially, this article does not endorse any specific position about AAs' metaphysical or moral status. It also does not insinuate that AAs have independent legal status: their normative impact on rights and duties of legal persons is intrinsically tied to the deployment relation as a whole—the deployment risk—implicating deployers' responsibility, not their own.

Part IV sketches the relation between pure legal agency and ordinary legal agency, which is constituted by two legal persons. Drawing on Lionel Smith's interpretation of fiduciary duties, it identifies a core idea of legal agency applicable to AAs: Legal agency reflects a social relation whereby one actor acts exclusively for the interests of another. It also highlights two fundamental aspects of legal agency that are absent in AA deployment. The first is that agency relations are typically formed by mutual assent of agent and principal. The second is that central to agency relations are fiduciary duties owed by agents to their principals.

The article concludes by offering a way to think about these doctrinal requirements in the context of AAs. It argues that mutual assent and fiduciary duties are necessary to constitute legal agency relations specifically where legal agents are also legal persons with their own subjective first-personal interests. Since AAs are not legal persons, but sophisticated instrumentalities, they inherently occupy a social role akin to legal agents: They are deployed to act exclusively for deployers' purposes, and lack subjective interests of their own, rendering doctrinal requirements of mutual assent and fiduciary duties redundant. While only a preliminary sketch, the pure legal agency account plausibly grounds deployers' vicarious liability as a matter of tort law, or so the argument goes.

Before continuing further, I must address, albeit briefly, the reasons for undertaking doctrinal analysis of AA-caused harm, and particularly, in the form of vicarious liability. One may reasonably object to this approach: The perspective of tort doctrine is too narrow to confront challenges posed by cutting-edge technology. It is preferable, the objection goes, to concentrate on identifying legitimate policy goals—*e.g.*, to incentivize innovation, establish consumer-safety product-design standards, deter cost-inefficient and unsafe product design, reduce or widely-spread accident costs, compensate innocent victims—and to craft regulatory schemes to promote them.[38] I concede that this criticism is sound: These policy objectives could be implemented through regulatory schemes to distribute fairly the benefits and burdens of AAs, including their accident costs. It is also possible that specific industries—*e.g.*, healthcare, transportation—will be subject to liability schemes that compensate accident victims.[39] The advantage of sidestepping tort law is clear: It avoids doctrines embodying notions of corrective justice and outcome responsibility, which have confusing application in the context of deploying emergent AAs which may cause accidental harm.

Nevertheless, there is reason to undertake common law tort analyses of AA-caused harm. First, tort analyses disclose legal options for policymakers to consider. Since legislative liability schemes may be modelled upon existing common law doctrines—though, perhaps, with modification—rigorous doctrinal analysis can reveal potential regulatory approaches, thereby contributing to the legislative process. Second, it is important to have a common law jurisprudence for AA-caused harm (alongside alternative liability schemes) if deployment of AAs becomes commonplace. Common law principles of general application will be necessary for cases of AA-caused harm that are not subject to specific regulatory schemes, and courts will look to tort law for a doctrinal resolution to the problem of AA-caused harm. Third, tort law embodies distinctive norms of interpersonal

38. For scholarly approaches that identify these policy concerns and consider tort and other regulatory schemes and standards to promote them, see *e.g.*, Mark A Geistfeld, "A Roadmap for Autonomous Vehicles: State Tort Liability, Automobile Insurance, and Federal Safety Regulations" (2017) 105 Calif L Rev 1611; Hubbard, *supra* note 22; Kenneth S Abraham & Robert L Rabin, "Automated Vehicles and Manufacturer Responsibility for Accidents: A New Legal Regime for a New Era" (2019) 105 Va L Rev 127.

39. For instance, Kenneth Abraham and Robert Rabin propose a new compensation regime, termed Manufacturers Enterprise Responsibility (MER), for victims of autonomous vehicles. MER would operate as a third-party insurance system funded by manufacturers of autonomous vehicles to compensate accident victims without resort to proof of negligence or product liability. It would be an exclusive remedy, replacing victims' rights to sue in tort. See Kenneth S Abraham & Robert L Rabin, "Automated Vehicles and Manufacturer Responsibility for Accidents: A New Legal Regime for a New Era" (2019) 105 Va L Rev 127.

responsibility, which provide a worthy contribution to scholarly work on the ethics of artificial intelligence and robotics. Tort doctrines and principles are normatively rich, offering insight into questions about risk, wrongdoing, and causation. Tort theory, then, contributes to ethical reflection about accountability for supplying and using emergent AAs that cause accidental harm. In addition to offering concrete doctrinal solutions, tort analyses should constitute part of the growing philosophical and ethical literature on artificial intelligence and robotics. Finally, tort analyses of AA-caused harm have self-standing value as an exercise in tort theory: to reflect upon tort law's applicability in the context of AA-mediated social relations. In contrast to the previous three reasons—relating to tort's contribution to resolving social and ethical problems of AA-caused harm—this fourth reason underscores that AA-caused harm is problematic *for tort law*. Tort law is concerned with a kind of interpersonal injustice—linking a plaintiff's loss to a particular defendant's tortious action—that may be simply irrelevant in the context of AA-mediated interaction. The concern is that emergent AAs—functionally independent, non-legal subjects—render tort law's internal normative viewpoint inapplicable. Tort analysis of AA-caused harm, in this respect, is important for its own sake: to study tort law, its proper scope, and its limits.

But these reasons take us only so far. The critic may still argue that this article's vicarious liability or pure legal agency approach is too extreme a reconception of tort law. Instead, the argument would go, it is preferable to make modest adjustments for which there is some historical precedent, such as reducing causation standards or plaintiffs' burden of proof, or embracing strict products liability.[40] My contention, nonetheless, is that any modification to tort law must address the fundamental issue of fair interpersonal conduct with respect to deploying AAs. Reduced-fault modifications, however, point toward alternative liability schemes; such fixes muddle the reasoned basis of agent-specific liability in tort. Absent notions of wrongful interpersonal conduct and outcome responsibility, tort's narrow focus on particular bilateral interactions between defendants and plaintiffs is unfair, preventing losses from being shared by all similarly situated enterprise participants benefiting from the relevant activity causing harm. These reduced-fault fixes also respond inadequately to the advent of emergent AAs, continuing to treat them as ordinary products. A doctrinal resolution is better served by thinking carefully about AAs' salient features—including their

40. These alternative modifications of tort are addressed, for instance, in Hubbard, *supra* note 22 at 1852.

extraordinary roles as functionally independent and rational instruments—that disrupt the structure of interpersonal rights and duties in tort.[41]

The pure legal agency or vicarious liability account is valuable for this reason: It offers a legal conception of AAs that captures their salient features and guides sound tort liability outcomes. My argument is that vicarious liability is a suitable doctrinal form to evaluate tort liability for AA-caused harm: It provides agent-specific grounds for liability specifically appropriate to the AA-deployment relation. My aim, in this sense, is to find coherence, in reflective equilibrium, between tort principles, the morality of corrective justice, and commonsensical liability outcomes for AA-caused harm.

## I. THE COMMON LAW OF VICARIOUS LIABILITY

Under the modern formulation of vicarious liability, employers are liable for torts committed by employees in the course of employment.[42] In this respect, vicarious liability has three doctrinal elements.[43] First, an employee *commits a tort*. The foundation of vicarious liability is the employee's tortious causation of harm. Liability is then broadened by imposing it on an additional defendant, the employer.[44] For this reason, an employer who is liable under vicarious liability typically has a right to indemnity from the employee, the tortious actor.[45]

Second, the person committing the tort is an *employee* of the defendant.[46] There is no conclusive test to determine whether a person is an employee rather than an independent contractor.[47] The central question is whether the person engaged to perform services does so as a servant for the employer, or whether the person acts independently "as a person in business on his own account."[48] Several factors assist this determination: the level of control the employer has over the worker's activities; whether the worker provides his own equipment or hires his own helpers; the degree of financial risk, investment and management held by

41. See Jack B Balkin, "The Path of Robotics Law" (2015) 6 Cal L Rev 45 at 47-48.
42. See Ernest J Weinrib, Tort Law: Cases and Materials, 4th ed (Emond Publishing, 2014) at 610 [Weinrib, Tort Law].
43. *Ibid*.
44. William Lloyd Prosser and Page Keeton, *Prosser and Keeton on Torts*, 5th (West Group, 1984) at 499.
45. *Romford Ice & Cold Storage Co v Lister*, [1957] AC 555 [*Lister*]; *London Drugs Ltd v Kuehne & Nagel International Ltd*, [1992] 3 SCR 299 [*London Drugs*].
46. See Weinrib, Tort Law, *supra* note 42 at 610.
47. *671122 Ontario Ltd v Sagaz Industries Canada Inc*, 2001 SCC 59 at paras 46-48.
48. *Ibid* at para 47.

the worker; and the worker's opportunity for profit in the activity.[49] There are also exceptional instances where vicarious liability extends to acts of independent contractors, such as where employers have non-delegable duties of care or where the work is especially dangerous.[50]

Third, the employee's tort is committed in *the course of employment*. In the doctrinal language, this requirement excludes torts committed while the employee is "on a frolic of his own," but includes torts committed while the defendant's "deviation from the prescribed task can be construed merely as a detour."[51] Even unauthorized or prohibited conduct can be construed as being in the course of employment if it is rightly regarded as a mode—albeit an improper one—of carrying out authorized acts of business.[52]

The function of vicarious liability is to hold employers responsible for tortious acts of employees that relate to characteristic risks of enterprise.[53] The critical analysis is whether an employee's tortious act is, in some sense, characteristic of and attributable to a risk created by the enterprise. The doctrinal significance of enterprise risks can be seen in the seminal case *Ira S Bushey*.[54] In that case, a drunk employee of the United States Cost Guard negligently turned wheels on a drydock wall, for no apparent reason, causing the ship to fall against, and damage, the drydock. Although the employee's tortious act did not have explicit purpose to serve the employer, the Court held the defendant government vicariously liable for the employee's negligent act. The Court stated that vicarious liability rests in a "deeply rooted sentiment" that a business enterprise should be responsible for tortious accidents that are characteristic of its activities.[55] An enterprise is liable for employees' tortious acts that relate to risks inherent in the ongoing activities of the enterprise. Importantly, however, since vicarious liability attaches to general ongoing risk of enterprise, an employee's particular tortious act need not be foreseeable to the employer as a real and substantial risk in the negligence sense. In *Ira S Bushey*, the Court found that the employee's negligent action—though not specifically foreseeable, nor aimed at a legitimate business rationale—was still part of the seafaring activity as it took place on the ship while attending to seafaring matters.[56] Since the seafaring enterprise employed seamen

49. *Ibid*.
50. See Weinrib, Tort Law, *supra* note 42 at 616.
51. *Ibid*.
52. Canadian Pacific Railway v Lockhart, [1941] SCR 278 at paras 9, 12.
53. *Ira S Bushey & Sons, Inc v United States*, 398 F 2d 167 (2nd Cir 1968) [*Bushey*].
54. *Ibid.*
55. *Ibid* at 171.
56. *Ibid* at 172.

who recurrently cross the drydock while drunk, potential tortious damage to the drydock was viewed as a risk inherent to the seafaring enterprise. By contrast, the Court noted, if the employee had damaged property on the street while walking to the ship, the incident would have related to his domestic life, not his seafaring activity, and liability would not extend to the enterprise itself.[57]

More recently, in a series of cases dealing with vicarious liability for sexual assault committed by employees, the Supreme Court of Canada (SCC) affirmed the centrality of enterprise risk to vicarious liability. In *Bazley v. Curry*, the Court held that a defendant non-profit organization operating a residential care facility was vicariously liable for its employee's sexual assault of children in its care.[58] Justice McLachlin, writing for the court, stated that employers are vicariously liable for unauthorized tortious acts of employees that fall within the ambit of risk that the enterprise creates or exacerbates.[59] The question of vicarious liability, then, concerns whether there is a sufficient nexus between the enterprise's risk and the employee's subsequent tort.[60] To ground vicarious liability, the enterprise needs to materially enhance the risk of tortious harm by putting the employee in a position conducive to preforming the tortious act.[61] This involves more than merely creating an opportunity for the tortious act, or being a but-for cause of the tortious act.[62] Yet, the enterprise risk does not need to be a foreseeable and substantial risk of the particular kind of resulting harm in the negligence sense. Enterprise risk is not tortious risk; it does not ground the tortious causation of harm itself. Rather, the doctrinal function of enterprise risk is to link employees' tortious causation of harm—whether negligence or intentional tort—to distinct risk inherent to the enterprise as a whole.[63] The determination of vicarious liability turns on judging the strength of the connection between employees' tortious acts and risks of enterprise. Vicarious liability captures instances where risks of enterprise could be seen to produce employees' tortious acts. In Justice McLachlin's words, "it must be possible to say that the employer *significantly* increased the risk of the harm by putting the employee in his or her position and requiring him to perform the assigned tasks."[64] The employment of the

57. *Ibid*.
58. *Bazley v Curry*, [1999] 2 SCR 534 [*Bazley*].
59. *Ibid* at para 37.
60. *Ibid*.
61. *Ibid* at para 40.
62. *Ibid*.
63. *Ibid* at para 39.
64. *Ibid* at para 42 [emphasis in original].

tortious actor must significantly increase the risk of the kind of tortious harm that ultimately ensues.

In *Bazley*, the defendant care facility was vicariously liable for the sexual assault committed by its employee, as it materially increased the risk of sexual assault. The defendant care facility provided its employee with opportunity for intimate private control—a parental relationship—over children in its care. The incidents of abuse were the "product of the special relationship of intimacy and respect the employer fostered, as well as the special opportunities for exploitation of that relationship it furnished."[65] By contrast, in *Jacobi v. Griffiths*, a companion case to *Bazley*, the SCC held that a defendant non-profit club providing group recreational activities for children was not vicariously liable for sexual assault committed by its program director.[66] In *Jacobi*, the opportunity that the club provided the program director for abuse was slight, as its activities were public, occurring in groups, and in the presence of volunteers.[67] The sexual abuse was possible only because the program director "subverted the public nature of the activities," enticing children to his home.[68] According to the Court, the director's tortious acts were insufficiently related to risks of enterprise, which merely provided the director opportunity to work with children in public and non-intimate settings. Likewise, several years later, in *E.B. v. Order of the Oblates*, the SCC held that a defendant residential school was not vicariously liable for sexual assault committed by its maintenance employee.[69] The Court found that the employee's duties did not include authority to be in situations of intimacy with students, nor to be alone with them.[70] While the residential school increased students' vulnerability in a general sense, there was no specific or enhanced risk due to employing the particular tort-committing-employee in his particular position to perform his assigned tasks.[71]

Accordingly, the critical point of analysis is whether an employee's tortious act can be traced to a distinct risk in employing the tortious actor. Vicarious liability reflects a particular kind of relation between the enterprise and the ensuing tortious act: The tortious act relates to a characteristic risk of enterprise in employing the tortious actor in a particular position to perform assigned

65. *Ibid* at para 58.
66. *Jacobi v Griffiths*, [1999] 2 SCR 570.
67. *Ibid* at para 80.
68. *Ibid*.
69. *EB v Order of the Oblates of Mary Immaculate in the Province of British Columbia*, 2005 SCC 60.
70. *Ibid* at para 48.
71. *Ibid*.

tasks. If this relation is present, unauthorized tortious acts are viewed as having been committed in the course of employment, and the employer is vicariously liable. Since the tortious act is committed in the course of employment, the act is deemed in law to have been committed on the employer's behalf, and the employer is seen to have (jointly) produced it. The tortious act is then legally attributable to the employer, grounding vicarious liability.

In this account, the notion of enterprise risk lies at the doctrinal core of vicarious liability. This is not to deny that there are also several policy goals purportedly advanced by vicarious liability.[72] The first is *compensation:* Plaintiffs gain compensatory access to an entity that is more likely to be financially capable of satisfying a judgment (*i.e.*, the employer enterprise with deeper pockets than its employees). The second is *deterrence*: Vicarious liability incentivizes employers to carefully select, train, and supervise employees, and to discipline employees who act wrongfully. The third is *enterprise liability*: Since enterprises advance their economic interests through assigning employment tasks to employees, it is fair to require enterprises to internalize the costs of torts committed by employees in the course of employment. The final policy rationale is *loss spreading*: Enterprises are generally in good position to obtain liability insurance and spread its costs to consumers through higher prices. Nevertheless, the imposition of vicarious liability still depends on a particular doctrinal configuration constituted by enterprise risk: In the absence of the right kind of relation between the enterprise risk and the ensuing tortious act, there is no vicarious liability, even if some of the policy goals could still be advanced.

In *Bazley*, the SCC emphasized that vicarious liability provides a *fair* means of compensating victims because the enterprise created the risk resulting in tortious conduct.[73] It also cautioned that if employees' tortious acts are not related to enterprise risk, imposing vicarious liability would wrongly reduce employers to being involuntary insurers, without suitably implicating the deterrence rationale.[74] In this way, the Court tied the compensation and deterrence policies to the doctrinal requirement of enterprise risk. The notion of enterprise risk is, then, the principled basis in tort to extend liability to employers for tortious harm committed by employees. It draws a link between the employee's tortious act and the employer's risk of employing the tortious actor to perform an assigned task. While vicarious liability may advance various social goals, such as compensation,

72. *London Drugs*, *supra* note 45, La Forest J.
73. *Bazley*, *supra* note 58 at paras 29-31.
74. *Ibid* at para 36.

deterrence, enterprise liability, or loss spreading, it only does so if the doctrinal requirement of enterprise risk is satisfied.

## II. VICARIOUS LIABILITY AND CORRECTIVE JUSTICE

As a practice of corrective justice, tort liability follows only if the plaintiff's loss is legally caused by a particular defendant's tortious act, grounding the defendant's agent-specific liability and duty to repair. The doctrine of vicarious liability appears, at first glance, to be inconsistent with this liability structure, as employers are liable for the unauthorized tortious acts of employees, without tortious conduct of their own. It is true that vicarious liability is still rooted in employees' actionable torts; it merely extends liability to employers where the torts are tied to risks of enterprise. However, this extension of liability to employers needs to be examined, as it obligates employers to repair plaintiffs' losses even though the risks of enterprise are not tortious, per se. Employers are then liable for losses they did not wrongfully cause. Perhaps, then, vicarious liability is an exception to the general principle that defendants' agent-specific liability attaches to their own tortious causation of harm.

There are three general ways to conceptualize employers' vicarious liability for unauthorized tortious acts committed by employees in the course of employment.[75] The first conception is that employers are burdened with the liability costs of tortious acts, though they are not the responsible actors. Vicarious liability, in this view, is a *re-distribution of liability costs* to employers who are, purportedly, the optimal cost-bearers.[76] It is not, however, a judgment about wrongful loss or outcome responsibility. As noted above, there are several policy rationales that partially explain the re-distribution of liability costs: compensation, deterrence, enterprise liability, and loss-spreading.

My contention, however, is that vicarious liability can be explained consistently with tort's structure of corrective justice. The key to make sense of this, arguably, is the doctrinal notion of enterprise risk; it suggests that vicarious liability responds to employers' special (agent-specific) responsibility

75. The three conceptions—whether vicarious liability is a mere distribution of liability costs to employers, or reflects a composite identity of employer-acting-through-employee, or finally, whether it reflects an attribution of the employee's tortious act to the employer, may be represented, respectively, by the approaches of Guido Calabresi, Ernest Weinrib, and Robert Stevens. See Guido Calabresi, "Some Thoughts on Risk Distribution and the Law of Torts" (1961) 70 Yale LJ 499, at 543-45; Weinrib, Private Law, *supra* note 30 at 186-87; and Robert Stevens, *Torts and Rights*, (Oxford University Press, 2007) at 259-62.

76. See *e.g.* Calabresi, *supra* note 75, at 543.

for the ensuing tort. Since it tracks enterprise risk, vicarious liability conveys that employers are, in some sense, outcome responsible for the tortious harm. This brings us to a second conception of vicarious liability: that the extension of liability to employers is actually a judgement about employers' tortious causation of harm.[77] Ernest Weinrib reasons, in this regard, that the agency relation between employers and employees affirms a legally constructed composite identity of employer-acting-through-employee: Since the employee is a mere cog in the enterprise, the employee's tortious act is an extended act of the enterprise as a whole.[78] Weinrib refers to this "composite" doer of harm as a "more inclusive legal persona," a legal fiction without empirical reality, reflecting law's construction of its own normative reality.[79] In this view, the employer is viewed as a tortious actor and liable as a matter corrective justice: Vicarious liability constructs a normative link between employer and victim, since the employee's tortious act is simply the act of the employer, as its principal. However, if an employee's tortious act is unauthorized, vicarious liability follows only if the proper relation between the tortious act and the enterprise is established, so that the employee is regarded as having acted *as part of the enterprise*.[80] This is the central role of enterprise risk: It establishes the requisite relation between the employee's tortious act and the enterprise—in which the employment itself produces risk of the particular tortious act—to attribute the tortious act to the enterprise despite its apparent non-authorization. If this relation is present, the employer is legally viewed as the tortious actor—that is, as having wrongfully caused the plaintiff's loss, grounding its agent-specific duty to repair.

However, there is a worthy counter-argument to this composite-doer theory: It largely collapses the distinct roles of employees and employers in causing tortious harm. The law distinguishes between their respective roles, as employers may seek indemnity from employees who cause unauthorized tortious harm in the course of business.[81] If employers are actually the tortious actors and employees' identities collapse as mere cogs in the enterprise, as the composite-doer conception suggests, it is illogical to legally empower employers to seek indemnity for their own tortious acts.

---

77. Weinrib, Private Law, *supra* note 30 at 186-87.
78. *Ibid*.
79. *Ibid*.
80. See Meir Dan-Cohen, "Responsibility and the Boundaries of the Self" (1992) 105 Harv L Rev 959 at 981-82.
81. *Lister*, *supra* note 45; *London Drugs*, *supra* note 45.

The assumption in this counter-argument is sound: As between employees and employers, the employees are the actual tortious actors. Employers may, therefore, seek indemnity from employees for unauthorized torts committed in the course of employment. At the same time, the employment relation has additional significance, in which employees act as extensions of the enterprise. This reflects a two-fold legal significance of employment relations. On the one hand, employees are independent legal actors, subjects of potential personal liability to victims or indemnity to employers for their tortious acts. On the other hand, employees act to advance their employers' purposes, not their own. Due to this latter feature, employees' tortious acts may be sufficiently tied to tasks of employment, and so attributable to the enterprise itself. This two-fold significance results in a multi-layered system of responsibility: As independent tortious actors, employees may be liable to indemnify employers for unauthorized torts committed in the course of employment. Simultaneously, employees' tortious acts are attributable to employers with respect to victims outside the employment relation.

This analysis suggests a third possible conception of vicarious liability, which features a subtle distinction between the *tortious action* and the *tortious actor*. While the employee is the tortious actor (the employer commits no initial tortious action), vicarious liability attributes the employee's tortious act to the employer vis-à-vis the plaintiff.[82] This framing gives rise to a further question: if the employer is not the actual tortious actor, why is the tortious action attributed to the employer? It relates, arguably, to a notion of collective or associational responsibility stemming from the social arrangement of employment: an employee's act also belongs to the employer when they act in a common pursuit.[83] This is the doctrinal role of enterprise risk: It links employees' unauthorized tortious acts to tasks of employment, and thereby, to the common pursuit of enterprise with which the employer is socially-identified. The employer

82. Robert Stevens maintains that vicarious liability involves an attribution of the employees' tortious *act* to the employer, not merely the *liability*. See Robert Stevens, *Torts and Rights*, (Oxford University Press, 2007) at 259-62.

83. See Dan-Cohen, *supra* note 80 at 985-86. As an illustration of collective responsibility, Meir Dan-Cohen offers the example of a team, where individual members use the first-person plural pronoun "we" to claim responsibility for acts of (other members of) the team, even if they did not play in the particular game. The collective team identity permits the attribution of acts by some individual members to other individual members. The attribution of the act—and responsibility for the act—exists by virtue of a shared identification with the collective team identity. This provides insight into vicarious liability: since the employer and employee share a common identity, defined by the pursuit of enterprise, the tortious actions of an employee—though acting alone and unauthorized—may be attributable to the employer where sufficiently linked to the pursuit of enterprise.

is then an additional individual responsibility base (*i.e.*, in addition to the employee) for tortious harm occurring in the pursuit of enterprise.

This third conception of vicarious liability—reflecting employers' identification with, and associational responsibility for, employees' tortious acts—is, in my view, consistent with tort's basic structure of agent-specific liability. Tort liability involves a normative attribution of a plaintiff's loss to a defendant's tortious act, as its legal cause. Vicarious liability involves this kind of normative attribution, but in way that is distinctly suitable to the employment context. It forms a particular legal construction of the employment relation: In cases of vicarious liability, employers and employees are seen to participate interdependently to produce tortious harm—that is, to jointly cause the tortious harm.[84] As a result, the law empowers plaintiffs with recourse against two defendants, employer and employee. In this way, the law recognizes the interdependence of employee and employer with respect to the commission of the tort: While the employee commits the completed tortious act, the potential for the tortious act was created by the employment itself. This does not mean the law construes the employer and employee as a single composite-doer-of-harm. Rather, the employer and employee are independent legal actors who jointly produce tortious harm, for which they are independent responsibility bases. Their joint production of tortious harm is not identical. While the tortious act is committed by the employee, it is also normatively attributable to the employer (vis-à-vis the plaintiff) due to the employer's self-identification with the employee's employment tasks, which comprise part of the collective goals of enterprise.[85]

The law, then, continues to view the employee as the tortious actor, and the employee is indeed a defendant in an immediate tort relation with the plaintiff, and also potentially required to indemnify the employer. However, the tortious action also belongs to the employer, since it occurred in the course of employment with which the employer is inescapably associated. In this view, vicarious liability is not a re-allocation of liability costs to promote external policy goals. It reflects an employer's *special responsibility for the wrongful loss*, due to its legal association and identification with the tortious action: The tortious causation of harm also belongs to the employer. This is a legal interpretation of the employment relation, where employers materially enhance the risk of employees committing specific torts by providing a particular employment opportunity. While the foundation of liability remains the employee's tortious act, liability is then extended to the

84. *Bazley*, *supra* note 58 at para 19.

85. On liability reflecting self-identification of an actor with a tortious action, see Dan-Cohen, *supra* note 80 at 981-86.

employer, who is responsible for associated acts of employees relating to the common pursuit of enterprise.

In this account, vicarious liability does not simply *re-distribute liability costs* of employees' tortious accidents to employers; rather, it *attributes employees' tortious acts* to employers. It does so by implicating the employment itself in the joint-production of tortious harm, drawing a normative link between the employee's tortious act and the risks of employment. If employment materially enhances the risk of the ensuing tort, the tortious act is sufficiently linked to the enterprise, despite being unauthorized. The tortious harm is then attributed to both components of the collectivity of enterprise, empowering the plaintiff to seek repair from two responsibility bases, employer and employee. Since the tortious act is also attributed to the enterprise itself, the employer could have an agent-specific obligation to repair the plaintiff's loss.

In sum, vicarious liability reflects a special legal relation between employers and employees. The employment relation recognizes the interdependence of employees and employers, as joint-producers of tortious harm while carrying out aims of enterprise. As independent legal actors, employees may be liable to victims (and employers) for torts committed in the course of employment. At the same time, vis-à-vis plaintiffs, employees' tortious acts are also legally attributable to employers. This layered liability outcome reflects employees' status as independent legal persons employed to act for the collective purposes of enterprise. It triggers liability consequences for employers who are, then, additional responsibility bases for tortious acts of the enterprise.

## III. VICARIOUS LIABILITY AND AA-CAUSED HARM

To review, the doctrinal problem of AA-caused harm reflects the problematic nature of AAs' social roles as mediators of human interaction. Emergent AAs are anticipated to perform tasks with a degree of functional independence. The normative significance of AAs' outward effects is not simply analogous to those of ordinary products. Since AAs are stimulated by environmental feedback, their effects are not necessarily traceable to human input. AA's outward effects are also not analogous to those of natural events, since AAs are deliberately deployed by humans for human purposes. AAs are, then, more like functional actors. As noted above, the laws of negligence and strict liability likely decline liability for AA-caused harm because users or distributors are not the relevant actors that cause the harm; AAs were deliberately deployed to perform the tasks instead. Since human distributors or users of AAs commit no initial tortious act, they are

not exposed to liability. These doctrines do not capture, as potential grounds for liability, the relevance of (reasonably) deploying another actor to perform a task.

The doctrinal form of vicarious liability, however, offers a promising solution to the problem of AA-caused harm, as it captures a form of associational responsibility for employing, or deploying, another actor to perform assigned tasks that increase risk of specific kinds of tortious harm. Victims of AA-caused harm should aim at this argument: Deployment of AAs involves characteristic risk of misalignment and AA-caused harm, analogous to human employment that provides special opportunity for employees to commit specific kinds of torts related to their tasks of employment. While deployment of AAs may not be unreasonably risky in the negligence sense, it involves characteristic risk of misalignment and AA-caused harm. It is comparable to employing human employees, which is not unreasonably risky in the negligence sense, but increases risks of tortious harm committed by employees. Notably, both human employees and AAs perform tasks to advance the goals of their employers and deployers, not their own. Moreover, both employers of human employees and deployers of AAs lack full operative control: Human employees can commit unauthorized tortious acts, and AAs' functional independence risks undesirable misalignment and AA-caused harm. Accordingly, the argument goes, just as employers are seen to jointly-produce their employees' tortious harm committed in the course of employment, so too should deployers of AAs be seen to jointly-produce AAs' tortious harm committed in the course of deployment.

This doctrinal solution is certainly controversial: it implicitly regards AAs as actors, rather than mere products. Critics may reasonably object to classifying inanimate artifacts as actors. In their view, automated vehicles or surgical robots are sophisticated tools, not more. Since AAs lack autonomy, intentional states, and consciousness, the argument goes, the analogy to human employees—legal and moral agents—is fictitious and potentially misleading. While employees' actions can give effect to legal consequences—for both themselves and employers—AAs do not *act* at all. To attribute legal consequences to AAs' *tortious actions*, then, is an artificial and misconceived fiction.

In truth, the assumption in this criticism is sound: In developing a theory of liability for AA-caused harm, erroneous characterizations of AAs must be avoided, such as attributions of mental states and consciousness. It is also crucial to be aware of the legal fiction that is involved in analogizing AAs to human employees. The ascription of agency to AAs is a legal metaphor, not a scientific or metaphysical description of their inner characteristics. Moreover, in selecting a liability category for AAs, it is best to not assume any particular reasonably-contested

philosophical view about sophisticated technologies and artificial intelligence, including positions in ongoing debates about whether technological artifacts can have autonomy or moral agency. The liability question is about legal duties that human (or corporate) actors owe to each other when supplying or using AAs. The liability determination should focus solely on the salient effects of AAs on human interaction, not on AAs' ontological or moral status.[86] The proper role of legal metaphor is limited: to evaluate AAs' normative position within legal relations—their impact on rights and duties of legal persons—not to conceptualize their general metaphysical character. This theory of vicarious liability for AA-caused harm, admittedly, adopts an intentional stance toward AAs, as legal actors, acting on behalf of human deployers.[87] However, this intentional stance is not an affirmation of a particular philosophical view of emergent technology. It is a legal and normative interpretation of the social relation between deployers, AAs, and their tort victims, where AAs are deployed to advance deployers' goals while producing characteristic risks of harm.

Aside from the theoretical leap to view AAs as legal actors, applying vicarious liability to AA-caused harm entails several doctrinal incongruencies. As stated above, vicarious liability typically involves three components: (a) a tort; (b) committed by an employee; (c) in the course of employment. Vicarious liability for AA-caused harm should then involve three analogous components: a *tort* committed by an AA, a *deployee*, in the *course of deployment*. There are several complications, however: Is there a way to differentiate between various instances of AA-caused harm as tortious or non-tortious? Can AAs and their deployers compose legal relations comparable to employment, thereby empowering AAs to give effect to legal consequences for deployers? Is there a way to establish whether AAs act within, or outside, the scope of deployment?

To impose vicarious liability in all instances of AA-caused harm would be, in effect, to impose an absolute liability standard, foreign to tort law. It is, therefore, necessary to discover a method to evaluate AA-caused harm as tortious or non-tortious. Establishing and defining the scope of legal agency relations between deployers and AAs, however, is especially problematic: If AAs are not legal persons, they presumably lack capacity to be party to legal relations. In this respect, the fact that human employees are legal persons has doctrinal consequences that may not be relevant in the context of AA-caused harm. First, under vicarious liability, employers may seek indemnity from employees. This is likely impossible in the context of AA-caused harm, unless legislation requires the

86. See generally Jack B Balkin, "The Path of Robotics Law" (2015) 6 Cal L Rev 45 at 47-48.
87. The intentional stance is discussed at Part III(A), below.

particular AA-tortfeasor to be insured. Moreover, where employees commit torts, plaintiffs are entitled to sue the employees directly. Again, this aspect of vicarious liability cannot apply to AA-caused harm, unless legislation deems the particular AA-tortfeasor to be a legal person and suable. These doctrinal divergences stem from the fact that, in this account, AAs have an ambiguous and problematic legal identity. On the one hand, AAs are deemed to be legal actors, as their outward behaviours may be evaluated as tortious, grounding deployers' vicarious liability. Yet, without legal personhood status, it is not clear that AAs can comprise legal agency relations and cause, by their own acts, legal effects for deployers.

It is not self-evident that there is a plausible solution to these complications. It is not surprising that emergent technologies do not fit seamlessly within a pre-existing framework for torts committed by human employees. In the remainder of this article, nonetheless, I attempt a preliminary sketch of a theory of vicarious liability for AA-caused harm. A successful theory needs to account for divergences between the ordinary doctrinal elements of vicarious liability and its application in the context of AA-caused harm. It should also make explicit the legal conception of AAs that is implicit in the vicarious liability account.

Let us begin by identifying the legal conception of AAs that is, in my view, implicit in the vicarious liability account. It involves recognizing a novel legal category for AAs: AAs are *pure* legal agents—that is, legal agents without legal personhood.[88] As legal agents, AAs have legal capacity to trigger legal consequences for deployers (who are legal persons). However, as entities without legal personhood, AAs do not produce legal consequences on their own account. The classification of AAs as legal agents without personhood can explain a difference between AAs and human employees in the way their respective agency relations would arise. AAs are deemed legal agents due to their inherent social role, as functional actors who act exclusively for deployers' purposes and interests, not their own. By contrast, in the employment context, since human employees are also independent legal persons with personal purposes and interests, agency relations arise due to particular (contractual) choices of employees to enter into employment. Furthermore, classifying AAs as legal agents without personhood helps to explain the various doctrinal divergences outlined above. In typical instances of vicarious liability, the tort is committed by a legal agent who is also

88. This category is proposed by Samir Chopra & Laurence F White, A Legal Theory for Autonomous Artificial Agents (University of Michigan Press, 2011) at 25. A pure legal agency classification also responds to Ryan Calo's suggestion that the law may need to adopt a "new category of legal subject, halfway between person and object." "Robotics and Lessons of Cyberlaw", *supra* note 3 at 549.

an independent legal person, grounding multiple responsibility bases; victims may sue either employees or employers, and employers may seek indemnity from employees. By contrast, in the context of AA-caused harm, since AAs are not legal persons, there is only one potential responsibility base: the deployers, not AAs themselves.

In sketching a theory of pure legal agency, the point of departure is tort law and vicarious liability. The legal conception of AAs that emerges in this account relates specifically to the role of legal agency in vicarious liability, based on an analogy between AA deployment and human employment. My aim is to make explicit a certain legal and normative interpretation of AA-mediated interaction that is implicit in the application of vicarious liability to AA-caused harm. I now turn to consider each component of vicarious liability for AA-caused harm: (a) tortious harm; (b) committed by an AA deployee; (c) in the course of deployment.

## A. AAS AS TORTFEASORS: ADOPTING AN INTENTIONAL STANCE

The foundation of vicarious liability for AA-caused harm is a completed tort committed by an AA. AAs are, then, viewed as potential tortfeasors, as bases of deployers' vicarious liability. The rough idea is this: The external actions of AAs are evaluated as tortious or non-tortious based on the degree of risk they pose to others. AAs' external actions—not their inner-algorithmic mechanisms—are evaluated as tortious. This approach resembles tort law's ordinary evaluation of tortious action. It considers whether particular external actions are consistent with systematic norms of rightful interaction, and an action is tortious if it poses a risk exceeding the level of risk typically assumed in ordinary patterns of interaction, as per an objective standard of care.[89] Tort law does not typically judge the inner mechanisms causing actions, such as actors' intentionality, nor does it judge the tortfeasors' subjective blameworthiness.[90] The same goes for evaluating AAs' actions under vicarious liability: AAs' external actions are tortious if they impose excessive levels of risk to other persons and property, inconsistent with norms of rightful interaction, as expected of the reasonable person. For example, if a self-driving vehicle fails to notice a stop sign and injures a pedestrian in an intersection, the self-driving vehicle will have acted negligently if, and only if, a reasonable person (driver) would have noticed the stop sign and avoided collision. In this way, the liability determination side-steps problematic and expensive inquiry into the inner defectiveness of AAs' algorithms, focussing

89. Stephen Perry, "Responsibility for Outcomes, Risk and the Law of Torts" in Gerald Postema, eds, *Philosophy and the Law of Torts* (Cambridge University Press, 2001) 72 at 111.
90. See Holmes, *supra* note 32.

instead on their outward behaviours and effects. It also thereby recognizes that AAs are not ordinary products; their sophisticated processes, outputs and effects resemble action.

Some legal scholars have proposed viewing AAs as tortfeasors. Ryan Abbott, for instance, reasons that if computers carry out activities once performed exclusively by humans and cause similar kinds of harm, they should be viewed as potential tortfeasors.[91] Abbott contrasts three kinds of cases that illustrate the relation between human action, technological artifacts, and applicable liability standards. In the first case, a human crane operator incorrectly identifies a drop off location and drops a steel frame on a passerby.[92] In this case, the operator's error is the cause of harm, and liability is judged under the law of negligence.[93] In the second case, a human crane operator manipulates a crane properly and under normal conditions, yet the crane is defective and tips over, landing on a passerby.[94] In this second case, the crane's defectiveness is the cause of harm, and liability is evaluated as a matter of products liability, not based on the operator's conduct.[95] In the third case, an unmanned autonomous computer operates a crane, misidentifies the drop off location and drops the steel frame on a passerby.[96] Abbott states that under prevailing tort law, scenarios two and three are treated alike; in both instances, harm is caused by defective products, and liability is evaluated under the products liability regime.[97] However, Abbott argues that scenario three—where the autonomous computer operator misidentifies the drop off location, dropping the frame on a passerby—is closely related to scenario one, where a human operator misidentifies the drop off location, dropping the frame on a passerby. Both scenarios involve the same kind of action and the same kind of physical result.[98] In Abbott's account, in the second case, the crane is a mere product, to be investigated for tortious defects. By contrast, in the third case, the "computer has stepped into the shoes of the worker; it has replaced a person, and it is performing in essentially the same manner as a person."[99] Therefore, Abbott reasons, in the third case, the autonomous crane should be viewed as a

91. Ryan Abbott, "The Reasonable Computer: Disrupting the Paradigm of Tort Liability" (2018) 86 Geo Wash L Rev at 23.
92. *Ibid* at 24.
93. *Ibid*.
94. *Ibid*.
95. *Ibid* at 25.
96. *Ibid*.
97. *Ibid*.
98. *Ibid*.
99. *Ibid*.

"computer tortfeasor," and manufacturers' liability should depend on whether the computer's external actions are tortious, not its product-defectiveness.[100] Notably, since Abbott grounds manufacturers' liability in the tortious actions of the machines they distribute, his theory is essentially a form of vicarious liability: He effectively deems computer tortfeasors to be agents of their manufacturers, thereby attributing computer torts to manufacturer principals.

Similarly, in the automated-vehicle context, Jeffrey Gurney proposes that the law treat manufacturers as drivers of automated vehicles they distribute, and that liability should be determined based on the vehicles' (un)reasonable actions, not its product-defects.[101] Gurney points out that automated vehicles provide little opportunity for driver control, especially when they are designed without steering wheels or breaks.[102] As a result, with automated vehicles, liability shifts from drivers to driving systems, falling mainly on manufacturers.[103] However, with respect to the theory of manufacturers' liability, Gurney suggests that the law treat manufacturers as the drivers of automated vehicles, turning liability into a "simple matter under negligence," rather than a "complicated matter under products liability."[104] Gurney warns that applying products liability to software and algorithmic defects is especially burdensome, not designed for everyday accidents.[105] Products liability litigation would be enormously expensive, as it involves complex and specialized evidence about algorithms and sensor data, requiring expert witnesses.[106] Gurney maintains that a reasonable driver standard can apply to automated vehicles.[107] Like human drivers, autonomous vehicles would be expected, for instance, to drive within proper lanes. What happens if, in a particular instance, an autonomous vehicle crosses the centre line causing a harmful accident? If a reasonable driver would not have crossed the center line, the action is deemed tortious and its manufacturer is liable.[108] Gurney's theory is vicarious liability in form: tortious acts of automated vehicles are attributed to their manufacturers. While Gurney does not explicitly acknowledge the parallel

100. *Ibid* at 26.
101. Jeffrey K Gurney, "Imputing Driverhood: Applying a Reasonable Driver Standard to Accidents Caused by Autonomous Vehicles" in Patrick Lin, Keith Abney & Ryan Jenkins, eds, *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence* (Oxford University Press, 2017) 51 at 59-60.
102. *Ibid* at 53.
103. *Ibid*.
104. *Ibid* at 59.
105. *Ibid* at 60.
106. *Ibid* at 55.
107. *Ibid* at 61.
108. *Ibid*.

to vicarious liability, his own phrase, "imputing driverhood," nicely captures the idea of agency, where *actions* of agents are attributed to their principals (not merely the resultant *liability costs of actions*, as elaborated upon above, in Part II).

In evaluating AAs' outward actions as tortious—rather than inspecting their inner-algorithmic-design for product defects—the law implicitly adopts a legal intentional stance toward AAs as legal actors.[109] In this respect, I allude to a central concept in intentional systems theory, formulated by Daniel Dennett.[110] In brief, Dennett sets out three basic stances used to interpret or predict the behaviour of other objects or entities. The first is the *physical stance*, where an object's physical constitution and the laws of physics are used to predict what an object will do. The physical stance is typically used, for instance, when predicting the outcome of releasing a stone from one's hand.[111] The second is the *design stance*, where an assumption is made that an object is designed to function in a particular way, and its behaviour is predicted based on this assumption.[112] The design stance is adopted with respect to objects like alarm clocks or chainsaws, where users typically assume that if they follow design instructions, the objects will operate as they are designed to function.[113] It is greatly impractical to predict or interpret the behaviour of these objects using a physical stance, to scrutinize their physical properties and the laws of physics to work out how to manipulate them effectively.[114] The third is the *intentional stance*, where an object is treated as an agent, with beliefs and desires, as well as the rationality to implement these imputed beliefs and desires.[115] According to Dennett, the intentional stance is useful where an object's behaviour is extremely complicated and most-easily predicted or interpreted by attributing to it a sense of rationality and goal-oriented behaviour.[116] For instance, to win a game of chess against a computer, it is best to predict the computer's moves as if it is a rational agent who knows the rules of chess and wants to win the game. This approach is more effective than inspecting its internal design, to calculate the many lines of computer code that determine

109. Chopra and White generally argue in favour of adopting an intentional stance towards AAs. See Chopra & White, *supra* note 88 at 11-17.
110. Daniel Dennett, "Intentional Systems Theory" in Ansgar Beckermann, Brian P McLaughlin & Sven Walter, eds, *The Oxford Handbook of Philosophy of Mind* (Oxford University Press, 2009) 339.
111. *Ibid* at 340.
112. *Ibid*.
113. *Ibid*.
114. *Ibid*.
115. *Ibid*.
116. *Ibid* at 340-41.

its next move.[117] As Dennett emphasizes, in instances like chess-playing, the intentional stance works because the computer is designed to "'reason' about the best move to make in the highly rationalistic setting of chess."[118] The intentional stance may then be appreciated as a sub-species of the design stance: The intentional stance works because the object is designed to behave rationally.[119]

Dennett also extends this key insight to human interaction: Humans view each other as intentional systems, using attributions of beliefs, desires, and rationality to anticipate or interpret human action, while remaining ignorant (for the most part) of peoples' actual internal mental processes.[120] In this view, the intentional stance is adopted with respect to a continuum of intentional systems, including humans, animals, and sophisticated artifacts—a range of instances where the attribution of intentional states gives meaning to behaviour where internal states are largely unknown.[121] Moreover, Dennett rejects the notion that, for sophisticated technological artifacts, the intentional stance is merely *derived intentionality*, but not *original intentionality*, or that it is *metaphorical*, not *literal*.[122] The point is that the intentional stance works to predict or interpret the object's behaviour regardless of whether the attributed goals are really appreciated, genuine or otherwise. There are still differences between various intentional systems; some are simple, others more complex.[123] Dennett distinguishes between first-order intentional systems—whose behaviour is predictable by attributing beliefs or desires to it—and second-order (or third or fourth, et cetera) systems, whose behaviour is predictable by attributing to it beliefs about beliefs (or desires). In Dennett's view, however, the intentional stance exploits the "deep similarity" between the whole continuum of intentional systems to investigate the differences between them.[124] The intentional stance is, then, a "theory-neutral way of capturing the cognitive competences of different organisms (or other agents) without committing the investigator to overspecific hypotheses about the internal structures that underlie the competences."[125]

Returning to tort law, to view AAs as tortfeasors is to assume an intentional stance toward AAs, conceived of as legal actors. The liability determination turns

117. *Ibid* at 341.
118. *Ibid*.
119. *Ibid*.
120. *Ibid* at 341-42.
121. *Ibid* at 342-43.
122. *Ibid* at 343.
123. *Ibid* at 344.
124. *Ibid*.
125. *Ibid*.

on whether their actions are consistent with the norms of rightful interaction. It does not turn on whether they are defective products due to tortious actions by other legal persons (*i.e.*, manufacturers or designers). To legally regard AAs' outward effects as actions—and subject to norms of rightful interaction—represents a legal move toward an intentional stance, away from a design stance. Typically, with respect to products, the law assumes a design stance: The product is expected to function according to its marketed design and operating instructions. If it does not, the law investigates whether it was designed, manufactured, or used tortiously by a responsible legal subject. Under a design stance, the law does not consider whether the product acted wrongfully, as the product does not act in the requisite legal sense. Rather, product defectiveness triggers an investigation of its legally responsible cause by an action of its manufacturer, designer, or user. By contrast, in applying vicarious liability to AA-caused harm, the law adopts an intentional stance: Tort analysis revolves around the reasonableness of AAs' external actions. The law elevates AAs' outward effects to the realm of legal action, with legal capacity to commit torts.

To be clear, in this context, to adopt an intentional stance is to assume AAs' outward effects are actions for the purposes of tort law, circumventing the need to further investigate the actions of manufacturers, designers, or users. Crucially, however, the legal intentional stance should not implicate any further philosophical positions, such as the general appropriateness of attributing mental states to sophisticated technologies, as under intentional systems theory. It would be a mistake for the law to take a position on any particular philosophy of mind. Accordingly, this vicarious liability account does not take a position on whether intentional systems theory is a suitable way to make sense of commonly used mentalistic terms such as beliefs, desires, or intentions. It also does not take a position on whether humans, animals, and technological artifacts represent a genuine continuum of related intentional systems. I emphasize, in this respect, that under standard accounts of agency, AAs are not truly agents, as originators of something that counts as action.[126] According to standard accounts, to count as actions, one's doings must be causally related to particular intentional mental

126. Kenneth Einar Himma, "Artificial Agency, Consciousness, and the Criteria for Moral Agency: What Properties Must an Artificial Agent Have To Be a Moral Agent?" (2009) 11 Ethics Info Tech 19 at 20-29.

states, such as a belief and desire pair or volition.[127] Crucially, if agency implicates intentional mental states, it also presupposes consciousness, as mental states are privately observable by introspection.[128] Pursuant to standard accounts of agency, therefore, absent mental states and consciousness, AAs cannot be agents at all.[129]

Nevertheless, intentional systems theory helps to illuminate the legal move to vicarious liability for AA-caused harm. It involves an altered *stance*: (1) to view AAs as rational actors that carry out goal-oriented tasks for human or corporate deployers; and (2) to evaluate AAs' actions according to an objective standard of rightful interaction. It is still crucial to acknowledge, however, that adopting an intentional stance is a legal fiction. If the outward effects of AAs are deemed to be actions for liability purposes, it is necessary to confront and reject the implicit supposition that AAs have (something comparable to) mental states causing their outward effects. After all, in ordinary instances, tort actions are necessarily expressions of volition, implicating a cluster of intentional states, such as intentions to act, beliefs, and desires (including second-order beliefs and desires).[130] To treat AAs as legal agents—elevating their outward effects to the realm of tortious action—problematically ensnares the possibility of artificial agency. I therefore emphasize that adopting an intentional stance toward AAs is a liability maneuver: AAs' agency is a legal fiction, not the legal recognition of a pre-existing metaphysical truth.[131]

---

127. *Ibid* at 20. Moreover, under standard accounts, moral agency—that is, to be subject to moral obligations and to be held accountable for one's actions—necessarily involves two further capacities. First, the capacity to freely act, which necessarily entails—under both libertarian and compatibilist conceptions—that actors (directly) cause their own behaviour and are not directly compelled by something external. Second, the capacity to engage in moral reasoning, which includes the ability to identify moral concepts and principles—to differentiate wrong from right—and apply them to specific contexts. *Ibid* at 21-24.
128. *Ibid* at 24-27.
129. *Ibid* at 28-29.
130. Restatement (Second) of Torts § 2 (1965).
131. I do not mean to use the term legal fiction in a pejorative sense. AAs' legal agency is a fiction in the sense that it does not necessarily correspond to some real metaphysical truth about artificial agency. However, as a legal fiction, it is a conceptual and normative interpretation of the relational structure of AA-mediated human interaction, which is, in my view, coherent and justified as a matter of tort law, as argued below in the text. As a general matter, since AA-caused harm demands reconstruction of tort doctrine, involving new forms of analogical reasoning, it is not surprising if it also involves making use of a legal fiction. On this point, see LL Fuller, "Legal Fictions" (1931) 25 Ill L Rev 513 at 527-28. Fuller observes that "[d]eveloping fields of law, fields where new social and business practices are necessitating a reconstruction of legal doctrine, nearly always present 'artificial construction,' and, in many cases, outright fictions."

However, this gives rise to the following questions: Is this legal fiction coherent and justified as a matter of tort law? Is it feasible to adopt an intentional stance toward AAs' outward effects, deemed to be tortious actions, without taking a position on the possibility of artificial agency? In my view, adopting this legal fiction is both intelligible and sensible, as a matter of tort law. First, under intentional systems theory, the move to an intentional stance is justified if the intentional stance enables better prediction or interpretation of an entity's behaviour than under the alternative physical or design stances. The liability maneuver to adopt an intentional stance follows similar reasoning, though with a legal spin: Liability for AAs' harmful effects is more easily addressed under an intentional stance—by evaluating their external behaviours—than under a design stance, entailing onerous scrutiny of AAs' defective algorithms. To insist on products liability for AA-caused harm is the legal equivalent of attempting to beat a chess-playing-computer by inspecting its internal design and calculating its many lines of code. Jeffrey Gurney emphasizes this point: "Treating the manufacturer as the driver makes what would have been a complicated matter under products liability a simple matter under negligence."[132] Accordingly, adopting an intentional stance toward AAs—as a legal fiction—serves a worthy and pragmatic legal purpose: to streamline the liability evaluation.

Second, tort law involves judgement about external actions, not actors' inner mental states or subjective blameworthiness. It may be conceded that since tort involves judgment of actions, the object of judgment is typically an expression of volition, implicating a mental state. Nevertheless, while the existence of a mental state is typically a pre-condition of tortious action, it is not the mental state itself that is judged as tortious. Tort liability does not respond to defendants' moral blameworthiness; it responds to external actions that are inconsistent with norms of rightful interaction, defined by an objective standard of care. Tort liability typically obligates defendants to repair losses caused by acts that are inconsistent with these norms, regardless of whether the actor could have acted differently in the particular circumstances. This approach can be coherently extended to the outward effects of AAs that are action-like, such as where a self-driving car fails to stop at a stop-sign and causes a harmful accident. Tort liability may assess whether their behaviours are consistent with objective standards of interaction expected of reasonable persons without declaring that AAs have, or do not have, intentional states.

Third, adopting an intentional stance toward AAs is arguably justified in the context of vicarious liability, which plainly attributes AAs' actions to deployers.

132. Gurney, *supra* note 101 at 59.

Since AAs are *pure* legal agents, the implicit ascription of intentionality to AAs actually belongs to their human or corporate deployers. If AAs are viewed as goal-oriented actors, it is really the deployers' goals that are referenced. AAs' legal agency, in this sense, encompasses a larger system: *AAs as extensions of deployers*. As pure legal agents, AAs' actions express deployers' intentional states, goals, and rationality—not their own. AAs' pure legal agency status is then a conceptual and normative interpretation of their deployers' extended actions: deployers-acting-through-their-deployees. In this respect, I retrieve the interpretation of vicarious liability where employer and employee are construed as a composite doer. In the context of AAs, this construction is particularly illuminating. Since AAs do not have independent legal standing, their intentional actions belong solely to their principals. AAs' outward effects are deemed to be intentional actions because AAs are designed and deployed to rationally achieve assigned tasks. AAs' actions should thus be viewed as extensions of deployers' intentionality. The evaluations of AAs' actions could then serve as surrogates for evaluating the responsibility of deployers.

In sum, viewing AAs as tortfeasors entails adopting an intentional stance in which AAs have legal capacity to commit tortious acts. This is a pragmatic liability maneuver with plausible theoretical foundations. Tort law can coherently evaluate AAs' outward effects as tortious actions without taking a position on whether they (can) have actual intentional states. After all, tort evaluations are judgments about external actions, not the intentional states that motivate them. Moreover, in the context of AAs, adopting an intentional stance is best understood as a legal interpretation of a larger system, consisting of both deployers and AAs: AAs' actions are extensions of deployers' intentionality. Treating AAs as tortfeasors is, then, intrinsically tied to their position within agency relations, as *deployees*, to which I turn next.

## B. AAS' DEPLOYMENT AS AN AGENCY RELATION

Vicarious liability applies to torts committed by employees in the course of employment. In the context of AA-caused harm, pursuant to an intentional stance, AAs may be deemed to be tortfeasors. However, to complete vicarious liability, AAs need to be part of legally recognized deployment relations, permitting attribution of AAs' tortious actions to deployers, analogous to employees' tortious acts committed in the course of employment. For obvious reasons, as mentioned above, AAs cannot enter legal agency relations by choice (in contrast to typical employment scenarios). Instead, in this vicarious liability account, the law would treat deployed AAs as inherent legal agents. Since AAs are deployed to achieve

their deployers' purposes, not their own, they inherently function in a way that is characteristic of legal agents. AAs occupy a unique middle-ground, as legal agents without personhood: They are instrumentally rational actors, but without their own purposes. This conception provides a theoretical foundation for the argument that AAs' tortious actions are extensions of deployers' intentionality, and therefore, a surrogate for evaluating the responsibility of deployers as a matter of vicarious liability.

To illustrate this legal conception, I turn to philosophy of technology literature dealing with the moral significance of technological artifacts. Broadly speaking, there are two important debates about the ascription of moral agency to sophisticated technological artifacts.[133] The first is the *autonomy debate*: whether artifacts are "mere instruments," their effects "fully explicable in terms of designer and user intentions," or whether artifacts may have a "degree of autonomy," actively causing effects in the world as goal-driven agents.[134] The second is the *moral relevance debate*: whether artifacts are necessarily "morally neutral means" to various human ends, or whether artifacts can be, in their own respect, "morally responsible agents."[135] As Christian Illies and Anthonie Meijers argue, the moral-relevance debate is largely dependent on the autonomy debate. Proponents of the moral neutrality position place "all moral weight on the intentionality of the users and designers," viewing artifacts as mere instruments of these human intentions.[136] Correspondingly, proponents of the position ascribing moral responsibility to artifacts also ascribe a degree of autonomy and agency to these artifacts, as active causes of morally significant effects.[137]

This framing admittedly oversimplifies the diversity of opinion about artificial moral agency; it is meant to capture the two poles—the most extreme positions—of the debate. My aim is to highlight specific formulations of technological artifacts' moral status, falling in between these two extremes, that illuminate the theoretical foundation of AAs' pure legal agency category. Deborah Johnson, for instance, argues that computer systems can be moral entities, but not independent, autonomous moral agents.[138] To begin, Johnson distinguishes between *artifacts*—as physical objects—and *technology*, which is a "combination

133. Christian Illies & Anthonie Meijers, "Artefacts Without Agency" (2009) 92 Monist 420.
134. *Ibid* at 421.
135. *Ibid*.
136. *Ibid*.
137. *Ibid*.
138. Deborah G Johnson, "Computer Systems: Moral Entities, But Not Moral Agents" (2006) 8 Ethics & IT 195 at 195.

of artifacts, social practices, social relationships, and systems of knowledge."[139] Johnson maintains that "artifacts are abstractions from reality:" To identify an artifact—as an independent object or entity—is to mentally separate it from the social context that gives it meaning and function.[140] Technological artifacts, such as computers and computer systems, exist due to, and as part of, complex systems of (human) social practices.[141] Johnson argues that the moral significance of computers cannot be conceived at "levels of abstraction that separate machine behaviour from the social practices of which it is a part and the humans who design and use it."[142] As Johnson forcefully reasons: "No matter how independently, automatically, and interactively computer systems of the future behave, they will be products (direct or indirect) of human behaviour, human social institutions, and human decision."[143]

Johnson also suggests that computer systems cannot be moral agents because they lack the capacity for voluntary behaviour that is essential to moral agency.[144] According to the standard formulation, voluntary intentional action entails an internal mental state causing an outward embodied event with an outward effect—harm or benefit—on a recipient of the action (a patient).[145] Accordingly, voluntary intentional behaviour involves a *reason explanation*, not only a causal explanation: Voluntary actions can be explained by mental states, such as beliefs, desires, and intentions to act.[146] Johnson emphasizes that computer system behaviour can occur where computers' internal states cause outward embodied events, with outward effects on patients.[147] Crucially, however, for computers, the inner state (causing outward effects) is not a mental state (*i.e.*, an intention to act), but some other mechanistic necessity.[148] Computer behaviour is amenable only to causal explanation, not reason explanation. Computer behaviour, then, does not entail free voluntary action and cannot ground moral agency.[149] Johnson acknowledges that machine learning techniques can enable computer behaviour that is, in her words, non-deterministic—*i.e.*, not directly linked to programmers'

139. *Ibid* at 197.
140. *Ibid*.
141. *Ibid*.
142. *Ibid* at 198.
143. *Ibid* at 197.
144. *Ibid* at 198-200.
145. *Ibid*.
146. *Ibid*.
147. *Ibid* at 199.
148. *Ibid*.
149. *Ibid*.

input—as in the case of neural networks.[150] However, she maintains that, "we have no way of knowing whether the non-deterministic character of human behaviour and non-deterministic behaviour of computer systems are or will be alike in the morally relevant (and admittedly mysterious) way."[151]

Johnson still maintains, however, that embodied computer systems are not morally neutral symbolic systems. Since they *behave* in the world—potentially producing effects on moral patients—mechanistic computer systems can have moral character, including intentionality (though without intentions to act).[152] According to Johnson, computer systems' intentionality relates to its programmed functions, "to behave in certain ways, given certain input," that is, to transform certain inputs into particular kinds of outputs.[153] Johnson argues that the intentionality of computer systems is connected to the intentionality of its designers and users.[154] When designing computer systems, designers "poise them to behave in certain ways;" the computer systems then "remain poised to behave in those ways."[155] Likewise, the intentionality of computer systems remains latent without user activation.[156] In this way, computer system intentionality is twofold: While dependent on initial human intentionality of designers and users, once initiated, computer systems independently produce certain intended states of affairs without further human intervention.[157] Johnson reasons that while computer systems—taken independently—are not moral agents, they are components of a broader moral analysis involving a triad of intentionality: that of users, designers, and computer systems.[158] Where humans use computer systems to achieve particular tasks, the computer systems do not act alone. The computer system is "part of an action but it is not alone an actor;" rather, the "triad of designer, artifact and user act(ed) as one."[159] Accordingly, while machines cannot be moral agents, they are moral entities. As components of human moral agency, they aid particular kinds of human action, and enable extended human intentionality and efficacy—that is, actions that are otherwise

150. *Ibid* at 200.
151. *Ibid*.
152. *Ibid* at 200-201.
153. *Ibid* at 201.
154. *Ibid*.
155. *Ibid*.
156. *Ibid*.
157. *Ibid* at 202.
158. *Ibid*.
159. *Ibid* at 203.

more difficult or impossible.[160] Johnson also argues that computer behaviour expresses designers' and users' intentionality even where designers and users are unable to predict precisely what the computer system will do.[161] In these instances, designers and users are simply engaging in risky behaviour, initiating actions with consequences that they cannot foresee. However, the computer behaviour still expresses intentionality and efficacy that its designers and users "put into the world."[162]

In a similar vein, Johnson and Thomas Powers recommend thinking about computers' moral agency as a kind of surrogate agency.[163] In standard accounts of agency, moral agents act from a first-person perspective, pursuing their own interests based on their own beliefs about the world.[164] By contrast, surrogate agents act (primarily) from a third-person perspective, pursuing the interests of their clients, not their own.[165] Johnson and Powers reason that computer systems' agency is akin to that of surrogate agents: "[They] are designed and deployed to do tasks assigned to them *by* humans."[166] Importantly, this argument does not depend on whether computer systems have actual autonomy, and largely bypasses the artificial intelligence debate. Rather, it is the link between computer systems and human interests that makes computer systems objects of moral evaluation.[167] Johnson and Powers reject the impulse to speak about computers' actions in psychological terms, as computer systems do not have their own first-person interests, nor a moral psychology.[168] Computer systems (exclusively) advance second-order interests: They represent users' interests and functionally perform tasks on their behalf.[169] Computer systems' second-order interests are not really their own; the second-order interest is simply a combination of the

160. *Ibid*.
161. *Ibid*.
162. *Ibid*.
163. Deborah G Johnson & Thomas M Powers, "Computers as Surrogate Agents" in J van den Hoven & J Weckert, eds, *Information Technology and Moral Philosophy* (Cambridge University Press, 2008) 251.
164. *Ibid* at 252.
165. *Ibid*. Surrogate agents—for example, lawyers, accountants, and executors—are typically constrained by systems of rules imposing duties to pursue the interests of their clients. Surrogate agents may commit wrongdoing by incompetence or by intentionally violating their duties by pursuing their own interests, rather than their clients' interests. *Ibid* at 252-54.
166. *Ibid* at 255 [emphasis in original].
167. *Ibid* at 258.
168. *Ibid*.
169. *Ibid* at 259.

computer program and user input.[170] This marks a significant contrast with human surrogate agents, who need to psychologically align their personal first-order interests with their second-order interests pursuant to the third-person perspective required for the job.[171]

Johnson and Powers emphasize that in evaluating moral responsibility for computer behaviour, it is a mistake to focus exclusively on human designers and users, ignoring the special role of computer systems which "constrain, facilitate and…shape what humans do."[172] At the same time, they deny that computer systems can be morally responsible since they do not have their own first-order interests, nor the moral psychology or freedom required under standard accounts of moral agency.[173] Johnson and Powers acknowledge that they have not precisely defined the instances where designers or users may—or may not—be responsible, liable, or rightly-blamed for behaviour of computer systems.[174] Nevertheless, their framing of computer systems—as surrogate agents and extensions of human intentionality—is valuable. It highlights a distinctive kind of normative relation between human deployers and sophisticated technologies they deploy—the kind of relation that the law could capture with a pure legal agency category.

Another notable view is the composite agency theory, also known as extended agency theory, formulated by F. Allan Hanson.[175] Hanson's theory begins with the uncontroversial idea that responsibility for a deed typically falls upon its doer.[176] However, where both human and nonhuman entities are necessary to accomplish a certain deed, Hanson argues that both human and non-human components comprise the relevant agency, "the doer of the deed."[177] At the core of composite agency theory is the contention that "humans do not and cannot act alone in order to accomplish what they do."[178] Rather, action is undertaken by "inter-related combinations of human and nonhuman elements."[179] The causal

---

170. *Ibid*.
171. *Ibid* at 260.
172. *Ibid* at 269.
173. *Ibid* at 270.
174. *Ibid* at 269.
175. F Allan Hanson, "Which Came First, the Doer or the Deed?" in Peter Kroes & Peter-Paul Verbeek, eds, *The Moral Status of Technical Artifacts* (Springer Science+Business Media, 2014) 55 [Hanson, "Doer or the Deed"].
176. *Ibid* at 56.
177. *Ibid* at 60.
178. *Ibid*.
179. F Allan Hanson, "The Anachronism of Moral Individualism and the Responsibility of Extended Agency" (2008) 7 Phenomenology & Cognitive Sci 415 at 416 [Hanson, "Anachronism"].

responsibility for an action's consequences, then, lies with the extended agency as a whole, which includes all necessary components of the action, both human and nonhuman.[180] In this view, the notion of agency is fluid, defined to include "the lines of communication essential to [the] activity," its components varying with the nature of the activity.[181] In composite agency theory, the doer is more precisely identified as a verb—"an embodied activity"—rather than a noun, as "a collection of objects".[182]

Hanson argues further: If action is explained in terms of composite agencies, the reduction of moral agency to human individuals is unwarranted.[183] Since moral responsibility lies with doers of deeds that have moral pertinence, the moral responsibility inquiry needs to isolate the relevant doer in any given situation. If the action is comprised of both human and nonhuman elements, the particular composite agency is the doer, the morally responsible agency.[184] According to Hanson, moral responsibility is not fundamentally distinct from causal responsibility: "It is a quality of causal responsibility that applies when the act has beneficial or detrimental consequences."[185] Where a composite agency acts with moral import—for example, causing harmful accident—the composite agency is morally responsible.[186]

Hanson's composite agency theory rejects the standard assumptions of methodological individualism, which reduces social behaviour to the actions of human individuals.[187] According to standard individualist accounts, machines, computers, and animals are mere objects that human agents manipulate in the course of their actions; they are not components of agency.[188] Methodological individualism predominantly rests on a modernist social frame: It conceives of the autonomous human individual as the most basic social unit, as an agent retaining a fixed identity while carrying out a variety of deeds over extended periods of time.[189] In this view, the human doer precedes the deed, and "remains stable as it moves from one deed to another."[190] By contrast, composite agency

180. *Ibid* at 418.
181. Hanson, "Doer or the Deed", *supra* note 175 at 61.
182. *Ibid*.
183. Hanson, "Anachronism", *supra* note 179 at 417.
184. Hanson, "Doer or the Deed", *supra* note 175 at 62.
185. Hanson, "Anachronism", *supra* note 179 at 418.
186. *Ibid*.
187. Hanson, "Doer or the Deed", *supra* note 175 at 56-57.
188. *Ibid*.
189. *Ibid* at 59.
190. *Ibid* at 61.

theory adopts a radically different account of agency, where "[t]he doer is defined by the deed."[191] In this view, there are no stable agents, only indefinite varieties of doers—that is, composite agencies—as defined by indefinite varieties of possible deeds.[192] Composite agency theory reflects a postmodern social frame, rejecting humanistic assumptions about the stable and centred autonomous nature of human agents.[193] It emphasizes the fluidity and indeterminacy of human agency, insisting that humans only exist in relation to the nonhuman, which underlies its embrace of an expanded ethical realm to include machine ethics.[194]

Hanson acknowledges that attributing moral responsibility to a composite agency as a whole is unconventional because nonhuman artifacts lack mental qualities such as awareness, intention, and foresight, which are indispensable to moral agency.[195] Nevertheless, Hanson argues that intentional acts are more precisely undertaken by, and attributable to, composite agencies consisting of both human and nonhuman components.[196] Hanson concedes that moral agency necessarily includes a human component with mental qualities, entailing the capacity for intelligent performance. Yet, he insists that an intelligent performance is still attributable to the composite agency as a whole, for it could not occur—nor could it even be intended—without contribution from nonhuman components as well.[197] After all, the possibility of particular actions is informed by particular means that are, or are not, available. In this sense, Hanson reasons that humans are changed by the technologies they use. For instance, he argues that "a-man-with-a-gun is a different being than the same man without a gun"; they are different moral subjects with distinct capabilities.[198] To be clear, Hanson does not claim that technological artifacts themselves can be morally responsible. He attributes moral responsibility to a composite agency as a whole; as the precise cause of actions and effects, the composite agency is the relevant moral subject.[199] The attribution of moral responsibility then extends to all components of a composite agency as joint responsibility. Moral responsibility necessarily extends to the human component, the locus of intentionality; however, the moral

191. *Ibid*.
192. *Ibid*.
193. *Ibid* at 65.
194. *Ibid* at 65-66.
195. *Ibid* at 63.
196. *Ibid* at 64.
197. *Ibid* at 65.
198. Hanson, "Anachronism", *supra* note 179 at 419.
199. F Allan Hanson, "Beyond the Skin Bag: On the Moral Responsibility of Extendent Agencies" (2009) 11 Ethics & Info Tech 91 at 95-96.

evaluation also encompasses nonhuman components that enable the particular intentional act (that is, as an act that could actually be intended).[200]

To return to tort law, these philosophical conceptions of technological artifacts as extending human agency can help construe the kind of normative relation between human deployers and AAs that is implicit in vicarious liability. As elaborated upon above in Part II, the kind of agent-specific responsibility that is captured by vicarious liability is explainable in two ways. First, it may reflect a conception of employer and employee as a composite doer: employer-acting-through-employee. Alternatively, it may involve the legal attribution of the employee's tortious act to the employer, reflecting the employer's associational responsibility for the act due to the employer's identification with the employee's acts committed in the course of employment, *i.e.*, in collective pursuit of enterprise. In the context of AA deployment, the composite doer conception is especially apt: Since AAs act solely for their deployers' interests—they do not have interests of their own—their actions may be viewed as legal extensions of their deployers' intentionality. Johnson makes this point by highlighting the "triad of intentionality"—of designers, users, and computer systems—as essentially intertwined components of a single moral analysis. Likewise, Hanson's conception of composite agency points to a robust conception of joint action and joint responsibility, where both human intentionality and AAs' functionality combine to cause morally significant effects such as accidental harm. The legal upshot is that judging AAs' actions as tortious must be part of a broader legal analysis linked to the deployment relation as a whole. The critical argument is this: The evaluation of AAs' actions as tortious—pursuant to a legal intentional stance—acquires its normative meaning from the deployer-AA relation. It is problematic to assess AAs' tortious actions independently—abstracted from the deployment relation—as this severs AAs' (tortious) actions from their source of intentionality. Accordingly, vicarious liability for AA-caused harm reflects a composite doer or associational conception of the deployment relation, where both human and AA components jointly produce tortious harm. The evaluation of AAs' actions as tortious can then serve as a surrogate for evaluating tortious responsibility of deployers.[201]

---

200. *Ibid* at 96-97.

201. This argument responds to Mark Chinen's suggestion that liability for AA-caused harm may need to be reframed in associative or collective terms. See Mark A Chinen, "The Co-Evolution of Autonomous Machines and Legal Responsibility" (2016) 20 Va JL & Tech 338 at 375-77. However, Chinen offers the model of associative responsibility as an alternative to tort law. The argument in the text is that vicarious liability embodies an ideal of associative responsibility as a matter of tort law.

This is certainly not a wholesale legal endorsement of Hanson's composite agency theory or Johnson's surrogate agency theory. These theories contain several features that should be distinguished from the issue of tort liability for AA-caused harm. First, they are about moral agency and moral responsibility, not legal agency and legal responsibility. Second, they apply to a broader class of technological artifacts, not only emergent AAs with functional autonomy. Third, Hanson's rejection of methodological individualism is probably inconsistent with the corrective justice account of tort, which rests on the possibility that responsibility for wrongful interaction is typically reducible to individual human agents. Finally, Hanson's conception of moral responsibility—which extends to all composite agencies—is a special kind of causal responsibility where acts have morally significant consequences, causing benefit or harm. In the corrective justice account, by contrast, tort liability involves wrongful causation of harm. Absent breach of duty, causation of loss is insufficient to ground liability, despite the moral significance of harm caused. Tort liability is not a practice of pure causal responsibility.

Nevertheless, these theories contain a key insight that is instructive in cases of AA-caused harm: The legal evaluation of AAs' actions must relate to the deployment relation as a whole, where AAs' tortious actions stand in as surrogates for deployers' actions. I stress that this is a *legal conception*, native to vicarious liability, a tort doctrine that captures a form of social interaction whereby a social actor is bound by the acts of an associated actor. In the context of AAs, vicarious liability reflects special responsibility for deploying AAs to perform particular tasks of deployment, producing the characteristic risk of misalignment and (tortious) AA-caused harm. I concede that adopting an intentional stance toward AAs is controversial. But it is justified in the context of *pure legal agency*, where the legal analysis maintains the form of vicarious liability, implicating a broader deployment relation (that includes human intentionality), not AAs' independent moral character.

### C. DEFINING THE SCOPE OF DEPLOYMENT

I have so far suggested that the deployment of AAs may be viewed as a legal agency relation between deployers and AAs, akin to an employment relation. The tortious acts of AAs committed in the course of deployment could then be attributed to deployers. However, this leads to a further question: How does one assess whether tortious actions of AAs fall within the scope of the deployment (agency) relation? If AAs' outward actions could fall outside scope of deployment, deployers would be able to disclaim the tortious acts of their AAs as unauthorized and frolic-like.

This potentially undermines the application of vicarious liability for AA-caused harm. David Vladeck cautions, in this respect, that if an autonomous machine acts in ways not pre-ordained by their initial programming—deciding for itself what course of action to take—the agency relation may breakdown.[202]

I aim to push back against Vladeck's view: Under vicarious liability, employers can be liable for *unauthorized* tortious acts of employees, provided such acts are sufficiently related to their assigned tasks of employment. Even prohibited conduct still falls within the course of employment where it can be regarded as a mode—albeit an improper one—of carrying out authorized acts of business.[203] Recall the influential case, *Ira S Bushey,* where an employee tortiously turned wheels on a drydock wall causing the ship to fall against the drydock, which had no legitimate business rationale.[204] Nevertheless, the Court found it to be part of the seafaring activity, as it took place on the ship while attending to seafaring matters.[205] In that case, the crucial factor grounding vicarious liability was that the risk of that kind of tortious damage to the drydock was characteristic of the seafaring enterprise, which employed seamen who recurrently crossed the drydock while drunk.[206] While the particular actions of the employee were not authorized business activities, they were sufficiently linked to the enterprise and the employment relation as a whole. Likewise, recall *Bazley v. Curry,* where the Supreme Court held the defendant care facility vicariously liable for its employee's sexual assault of children in its care.[207] The employee was certainly not authorized to commit intentional torts such as sexual assault. Moreover, the employee's sexual assault impeded, rather than advanced, the actual purposes of enterprise. Nevertheless, for the purposes of vicarious liability, the scope of employment is construed more broadly: If the particular assigned tasks of employment provide special opportunity for the particular kind of ensuing tortious harm, the tortious act falls within the course of employment so is attributable to the employer.[208]

Accordingly, deployers can be vicariously liable for tortious harm caused by AAs, even if AAs' actions are unauthorized and unwanted. At this early point, lacking experience with AA deployment, it is difficult to precisely define the scope of deployment. However, I will offer an initial way to think about this issue. The scope of an AA deployment relation should be defined by an AA's task: Deployers

202. *Supra* note 6 at 122-23.
203. *Bushey*, *supra* note 53.
204. *Ibid*.
205. *Ibid* at 172.
206. *Ibid* at 172.
207. *Bazley*, *supra* note 58.
208. *Ibid* at paras 37-42.

are vicariously liable where tortious harm caused by an AA falls within the scope of the characteristic risk of its deployment. The scope of AA deployment, then, is defined by the deployment risk, analogous to the doctrinal requirement of enterprise risk. As discussed, the notion of enterprise risk draws a link between an employee's tortious act and the risk in employing the tortious actor to perform an assigned task. Likewise, to ground vicarious liability, an AA's tortious act needs to be sufficiently linked to distinct risk in deploying the AA to perform its assigned task. Where there is a sufficient nexus between the resulting tortious harm and the deployment's characteristic risk—*i.e.*, the deployment risk—the deployment is properly implicated as jointly producing the tortious AA-caused harm. For example, in the context of automated vehicles, deployers may be vicariously liable for harm caused by AAs' driving-related errors. Deployers may not be vicariously liable, however, for accidents that fall outside the usual scope of driving. For instance, if a car is programmed to experiment with energy efficiency and starts its engine in a garage to recharge its battery causing a passenger's death by carbon monoxide, the resulting harm may fall outside the scope of driving-related harms.[209] The point is that vicarious liability for AA-caused harm takes this form: Deployers are liable where AAs cause tortious harm falling within the scope of risk associated with their ordinary tasks. The resulting AA-caused harm must correspond to a characteristic risk in deploying the particular AA, thereby conforming to the doctrinal structure of deployment risk (analogous to enterprise risk).

## D. WHO ARE THE DEPLOYERS?

The account sketched thus far has not addressed who the deployers are for the purposes of vicarious liability. Are they the users (or owners) of AAs or the distributors (*e.g.*, manufacturers, designers, retailers) of AAs?

One argument is that users should be vicariously liable, not distributors. Since AAs perform tasks for, and pursue the goals of, their users, they may be conceived of as users' legal agents. By contrast, AAs do not perform specific tasks for manufactures, designers, or retailers. The employment analogy, then, pertains more straightforwardly to AAs' users than to their distributors. The connection between AAs and their distributors, moreover, reflects a design stance. AAs are designed, manufactured, and distributed as products. It is only once users deploy AAs to perform tasks in social environments that AAs resemble social actors.

209. See Ryan Calo, "Robots as Legal Metaphors" (2016) 30 Harv JL & Tech 209 at 230.

Therefore, the argument concludes, AAs' tortious actions should belong to their users, not their distributors.

But there is also an argument to support imposing vicarious liability on AAs' distributors. AAs may be conceived of as agents of distributors since they reflect the intentionality that their designers and manufacturers build into them. As Deborah Johnson puts it, designers "poise them to behave in certain ways," and they "remain poised to behave in those ways."[210] Conceiving of AAs as legal agents of distributors also reflects their distinctive middle-ground character, falling somewhere between full-fledged *persons* and mere *things*.[211] AAs are artifacts; they are designed, produced and distributed as products. Products liability is, therefore, a conventional liability category for AA-caused harm, implicating distributors, not users. At the same time, AAs' emergent capabilities resist application of conventional products liability, which entails burdensome inspection of algorithms for defects, and misapprehends AAs' social and normative position as functionally independent instrumentalities. For these reasons, vicarious liability is an attractive solution. This framing points to a conception of AAs as sophisticated products with intentionality linked to their designers and manufacturers. Distributors should then be vicariously liable for AAs' tortious harm, as a sub-species of products liability: The vicarious liability standard reflects the fact that these products are emergent and act with functional independence. Moreover, regarding AAs as agents of manufacturers, rather than users, represents a more incremental development of tort law. As a sub-species of products liability, manufacturers' vicarious liability can be framed as a revamped consumer-expectations test: Consumers expect AAs to operate as reasonable persons would.[212] AAs' product defectiveness would then be determined with respect to their outward behaviour, not their internal design, as per the legal intentional stance.

I conclude this discussion without taking a position on this issue; it suffices to flag both approaches as possible tort doctrines. My reluctance to determine the identity of the deployer does not undermine this article's vicarious liability

---

210. Johnson, *supra* note 138 at 201.
211. Recall Ryan Calo's suggestion that the law may adopt a "new category of legal subject, halfway between person and object." Calo, "Robotics and Lessons of Cyberlaw", *supra* note 3 at 549.
212. Bryant Walker Smith also suggests that an automated driving system could be deemed defective under a consumer-expectations test if it does not perform as a reasonable driver would. See Smith, *supra* note 26 at 46. According to the argument in the text, Smith's application of the consumer-expectations test is actually a version of vicarious liability, as applied to autonomous machines.

analysis. My intent is to present vicarious liability as a suitable *form* of tort liability for AA-caused harm. The issue of which entities should be deployers for the purposes of vicarious liability—whether users or distributors—is a separate inquiry. It presupposes, however, an initial foundation that vicarious liability is a suitable form of liability for AA-caused harm. The aim of this article is to explore this foundation: Vicarious liability for AA-caused harm involves conceiving of AAs as pure legal agents with capacity to effect liability consequences for deployers. Within this account, however, either users or distributors could be deemed, in law, to be deployers for the purposes of vicarious liability. The choice between users and distributors, crucially, should be informed by more robust considerations of policy, a line of inquiry falling outside the scope of this article. The pure legal agency or vicarious liability account, in this sense, is underdetermined, so it can and should be supplemented with more concrete reasons relating to efficient accident-cost allocation, incentivizing technological innovation, securing compensatory access for faultless victims, and interpersonal accountability—on economic and ethical grounds.

## IV. TOWARD A THEORY OF AAS' PURE LEGAL AGENCY

In this study, the point of departure was tort law. The pure legal agency conception of AAs relates specifically to a theory of vicarious liability for AA-caused harm, based on an analogy between AA deployment and human employment. AAs are conceived of as pure legal agents, empowered to trigger their deployers' vicarious liability. Since this classification views AAs as legal agents, it also implicates the laws of agency more generally. AAs' pure legal agency category, then, needs to be situated within a broader agency framework, compared and contrasted with ordinary instances of legal agency constituted by two legal persons. In particular, I note two essential features that typically constitute legal agency relations: the first is a mutually manifested assent by principal and agent; the second is a set of fiduciary duties owed by agents to principals. A theory of AAs' pure legal agency must account for the apparent absence of these features. This section sketches the relation between pure legal agency and ordinary legal agency constituted by two legal persons. This account is admittedly partial and incomplete, however; it is only a preliminary point of entry toward a theory of pure legal agency.

Legal agency relations ordinarily arise upon mutually manifested assent by a principal and agent; that the agent acts for, and subject to the control of, the

principal.[213] The requisite manifestations of assent can occur in several ways: by contract, whether express or implied, oral or written; ratification, where the principal assents to an act after-the-fact; estoppel, where a person acts for another to an extent that causes others to reasonably believe an agency relation exists; or necessity, where a person acts for another in an emergency.[214] Samir Chopra and Laurence White note that legal agency relations may arise due to observable actions of principal and agent, without express agreement of agency.[215] This occurs, for instance, in situations of estoppel. AAs' pure legal agency would fall under this rubric; their legal agency status is inferred from their roles and behaviours. The initiation of an AA agency relation certainly diverges from typical instances of legal agency wherein legal agents are also legal persons who manifest assent in more obvious ways. Nevertheless, as Chopra and White maintain, there is a doctrinal basis for inferring AAs' legal agency from their deployment itself.[216]

A second fundamental feature of legal agency is the set of duties that agents owe to their principals. These include, among others, duties to obey principals' instructions, to act with skill and loyalty, and to protect confidential information.[217] We must ask: Can AAs have, and fulfil, duties to their deployers? Chopra and White argue that "artificial agents can be coherently understood as having duties to their principals if we can understand them as acting in conformity with statements that are best understood as their obligations to their principals."[218] This position reflects an intentional stance; AAs' performances of required tasks—for instance, submitting daily reports of earnings to principals—is interpreted as compliance with duties.[219]

My contention, however, is that the terminology of rights and duties—that AAs owe duties to deployers who hold correlative rights—is unfitting in the context of AAs' pure legal agency. Since AAs lack legal personhood, they probably cannot participate in juridical relations constituted by correlative rights and duties. If this is true, one may object to AAs' classification as legal agents; after all, fiduciary duties are central to legal agency. Nevertheless, my thought is that AAs' legal agency status—that is, their powers to effect legal consequences for deployers—is coherent even without owing fiduciary duties to deployers.

213. *Restatement (Third) of Agency* § 1.01 (2006).
214. Chopra & White, *supra* note 88 at 19, citing Robert W Emerson & John W Hardwicke, *Business Law*, 3rd ed (Barron's Educational Series, 1997) at 251.
215. Chopra & White, *supra* note 88 at 19-20.
216. *Ibid* at 20, n 34.
217. *Ibid* at 20.
218. *Ibid* at 21.
219. *Ibid*.

The rough argument is this: Fiduciary duties are required to constitute legal agency relations specifically where legal agents are independent legal persons with their own interests. In such cases, agents' conflicts of interest are a real concern, so fiduciary duties serve to ensure agents act only for the interests of their principals. By contrast, since AAs are *pure* legal agents, without interests of their own, fiduciary duties are largely redundant. AAs inherently and exclusively pursue their deployers' purposes.[220]

As a general matter, fiduciary obligations lie at the core of legal agency, integral to the exercise of fiduciary power. Fiduciary obligations are about ensuring the agent's loyalty: allegiance or dedication to the principal's cause rather than to self-interest.[221] Fiduciary relationships arise where a legal actor holds discretionary power: the capacity to exercise judgment for a beneficiary.[222] Lionel Smith understands the requirement of loyalty as a required manner of exercising judgment, where fiduciaries need to subjectively believe that their choices are in the best interests of beneficiaries.[223] While there are several constraints on fiduciaries' powers that are assessed objectively—such as duties of care, skill, and diligence—the requirement of loyalty has a subjective character.[224]

The fiduciary power and its requirement of loyalty are not simply contractual.[225] The grant of fiduciary power can be voluntary, as in typical principal-agent relationships; however, fiduciary powers may also exist as a matter of law, as in cases of company directors and executors who hold powers in managerial capacities.[226] The fiduciary requirement of loyalty is imposed by law if, and only if, a certain kind of role is assumed: "When one person acquires the authority to make decisions *on behalf* of another person, there is a partial transfer of autonomy."[227] In this situation, the fiduciary is authorized to act *for* the other, not just in a way that affects the other.[228] Smith argues that the requirement of loyalty is an inherent feature of the fiduciary relation: It is built into the power itself.[229] In fiduciary relations, he reasons, fiduciaries are empowered to

220. See Johnson & Powers, *supra* note 163 at 255-59; Johnson, *supra* note 138 at 201-203.
221. Lionel Smith, "Fiduciary Relationships: Ensuring the Loyal Exercise of Judgment on Behalf of Another" (2014) 130 Law Q Rev 608 at 608-609.
222. *Ibid* at 610.
223. *Ibid* at 611-12.
224. *Ibid* at 612.
225. *Ibid* at 613.
226. *Ibid* at 616.
227. *Ibid* at 613 [emphasis in original].
228. *Ibid*.
229. *Ibid* at 614.

exercise a part of the beneficiaries' autonomy by making decisions that belong to beneficiaries.[230] The fiduciary makes decisions *for* the beneficiary; that is, the fiduciary exercises the beneficiary's decision-making power.[231] Since the fiduciary power is about exercising the autonomy of another, the requirement of loyalty is not an external duty but an aspect of the power itself, the very meaning of acting *for* the beneficiary. The requirement of loyalty, then, is not a duty in the strict sense; rather, it is a power-conferring rule, with built-in limits to exercise the power loyally.[232] For this reason, the primary remedy for breaches of loyalty is rescission. This is analogous to situations where one enters into compelled agreements: A non-loyal exercise of judgment *for* the beneficiary is fundamentally flawed and incoherent, and prompts rescission.[233]

The fiduciary power and its requirement of loyalty are tied to two further rules: the no-conflicts rule, and the no-profits rule. The no-conflicts rule is as an extension of the loyalty requirement. In situations of conflict, it is impossible to be certain that extraneous considerations—*i.e.*, those negatively affecting loyal judgment—have been excluded.[234] The no-conflicts rule relates to the subjective character of loyalty; compliance cannot be objectively determined by evaluating the fiduciary's decision itself.[235] In situations of conflict, loyalty cannot be assured, so fiduciaries' legal acts are voidable by beneficiaries.[236] The no-profits rule, finally, permits beneficiaries to strip profits from fiduciaries gained through their fiduciary positions. According to Smith, recovery of profits is a rule of primary attribution, not a secondary rule resulting from fiduciary wrongdoing; it arises from the fiduciary relation itself, in which the fiduciary receives all profits *for* the beneficiary.[237] Again, this relates to the transfer of the beneficiary's autonomy to the fiduciary: Where the fiduciary gains something through the fiduciary position—while acting for the beneficiary—the gain is credited to the beneficiary as a primary right.[238]

Accordingly, the essential feature of the fiduciary relation is the transfer of discretionary decision-making power to the fiduciary, along with the fiduciary's subjective identification with the interests of the beneficiary, as represented by the

230. *Ibid.*
231. *Ibid.*
232. *Ibid* at 621.
233. *Ibid* at 620.
234. *Ibid* at 624.
235. *Ibid.*
236. *Ibid* at 625.
237. *Ibid* at 628.
238. *Ibid.*

requirement of loyalty. Fiduciary duties are intelligible in light of this juridical relation: They ensure that the fiduciary exercises discretionary decision-making power *for* the beneficiary, so that the fiduciary's decision can rightly be seen as the beneficiary's decision. Therefore, if the agency relation—as a fiduciary relation—has a core meaning, it is the transfer of discretionary decision-making power itself—*i.e.*, the power to act *for* another—not the fiduciary duties which support it.

Let us return to the pure legal agency account. AAs perform tasks for deployers with functional independence. In this sense, AAs' social role is analogous to that of legal agents: AAs are instrumentally rational actors who act exclusively for human or corporate purposes. They do not have their own subjective interests that need to be restrained. Their legal agency status is then intelligible without being constituted by fiduciary duties. In ordinary instances of legal agency, where agents are legal persons with their own personal interests, fiduciary duties are necessary to constitute agency relations. In these instances, absent fiduciary duties, there is no assurance that agents will act exclusively for their principals, subjectively identifying with their principals' interests. However, since AAs do not have subjective interests of their own, fiduciary duties are redundant.

Admittedly, this is not a comprehensive treatment of the laws of agency. To reiterate, in this article, the point of departure was tort law, and AAs' pure legal agency status relates specifically to a theory of vicarious liability for AA-caused harm, based on an analogy between AA deployment and human employment. The broader implications of AAs' pure legal agency status fall beyond the scope of this article. However, Smith's conception of legal agency as a transfer of discretionary decision-making power—the power to act *for* another—suggests that agency has a core meaning that is applicable to AAs. This core insight offers a plausible theoretical foundation to ground vicarious liability for AA-caused harm and is a point of entry toward a theory of pure legal agency.

## V. CONCLUSION

The doctrinal form of vicarious liability offers a promising basis to ground tort liability for AA-caused harm, upon which deployers are liable for tortious harm caused by AAs in the course of deployment. In this view, AAs' outward effects are evaluated as tortious (or non-tortious), pursuant to a legal intentional stance, alleviating the inefficient task of scrutinizing AAs' emergent algorithms to trace AAs' harmful effects to tortious human agency. AAs' tortious acts are then attributed to deployers when committed in the course of deployment—that is, when related to characteristic deployment risk. In this way, AAs resemble legal

agents, empowered to trigger liability consequences for deployers. In particular, AAs are *pure* legal agents without legal personhood.

AAs' pure legal agency is coherent and sensible in the context of vicarious liability, which implicates a broader deployment relation, and not AAs' intrinsic legal or moral character. The pure legal agency classification captures AAs' normative position within human relations as functionally independent and rational instruments deployed to act exclusively for human purposes. Evaluating AAs' external behaviours as tortious is a surrogate for determining their deployers' agent-specific (vicarious) liability as a matter of tort law. The pure legal agency classification, in this respect, offers a valuable strategy to promote pragmatic liability outcomes for AA-caused harm, in accordance with tort's doctrinal and theoretical structure of corrective justice.