MITCHELL | HAMLINE School of Law

Mitchell Hamline Law Review

Volume 46 | Issue 5 Article 2

2020

AI Entities as AI Agents: Artificial Intelligence Liability and the AI Respondeat Superior Analogy

Anat Lior

AI ENTITIES AS AI AGENTS: ARTIFICIAL INTELLIGENCE LIABILITY AND THE AI RESPONDEAT SUPERIOR ANALOGY

Anat Lior*

ABSTRACT

Artificial Intelligence (AI) based entities are already causing damages and fatalities in today's commercial world. As a result, the dispute about tort liability of AI-based machines, algorithms, agents, and robots is exponentially advancing in the scholarly world and outside of it. When it comes to AI accidents, different scholars and key figures in the AI industry advocate for different liability regimes. This ever-growing disagreement is condemning this new emergent technology, soon to be found in almost every home and street in the US and around the world, into a realm of regulatory uncertainty. This obstructs our ability to fully enjoy the many benefits AI has to offer us as consumers and as a society.

This Article advocates for the adoption and application of a strict liability regime on current and future AI accidents. It does so by delving into and exploring the realm of legal analogies in the AI context and promoting the agency analogy, and subsequently, the respondeat superior doctrine. This Article explains and justifies why the agency analogy is the best-suited one in contrast to other analogies which have been suggested in the context of AI liability (e.g., products, animals, electronic persons and even slaves). As a result, the intuitive application of the respondeat superior doctrine provides the AI industry with a much-needed underlying liability regime which will enable it to continue to evolve in the years to come, and its victims to receive remedy once accidents occur.

I. INTRODUCTION

The conundrum of Artificial Intelligence (AI) liability has become a real-life problem in society today. Scenarios that were once only plausible in science fiction novels and movies[1] are now transfusing themselves into our routine lives, causing manifold legal and social challenges. We are still far from the ominous scenario in which the machines take over humanity.[2] Nonetheless, today, AI-based robots and algorithms can and do inflict physical and non-physical damages upon us as a society and as individuals, while the legal approach to handling these damages is highly disputed.[3]

---

*J.S.D. Candidate at Yale Law School, Resident Fellow with the School's Information Society Project. I would like to deeply thank Professor Jack Balkin for his guidance and help through the different stages of this article; Professors Guido Calabresi and Daniel Markovitz for their comments; as well as participants at the Yale Information Society Project Fellows Writing Workshop; the Yale Law School J.S.D. Colloquium Workshop; and the 23rd UBC Interdisciplinary Legal Studies Graduate Innovation in Law and Policy Conference.

[1] *See, e.g.*, ISAAC ASIMOV, I, ROBOT (1950) (detailing fictional history of robotics); 2001: A SPACE ODYSSEY (Metro-Goldwyn-Mayer 1968) (imaging a world in which a space ship computer system, HAL, rebels against human astronauts); THE MATRIX (Warner Bros. 1999) (describing a world in which an artificial intelligence character, known as Agent Smith, keeps order in Matrix system by terminating troublesome programs and humans); THE TERMINATOR (Hemdale 1984) (envisioning a cyborg assassin, known as the Terminator, sent from the future by self-aware artificial intelligence program Skynet); HER (Annapurna Pictures 2013) (imaging a scenario where a man falls in love with Samantha, his operating system).

[2] *See, e.g.*, NICK BOSTROM, SUPERINTELLIGENCE: PATHS, DANGEROUS, STRATEGIES (2014) (arguing that if machine brains surpass human brains in general intelligence, this new superintelligence could replace humans as the dominant lifeform on Earth).

[3] *See* Jack M. Balkin, *The Path of Robotics Law*, 6 CALIF. L. REV. CIR. 45 (2015) (describing key problems that robotics present for law and exploring solutions). *See generally*, UGO PAGALLO, THE LAWS OF ROBOTS: CRIMES, CONTRACTS, AND TORTS (2013) (exploring how the design, construction, and use of robotics technology may affect today's legal systems and matters of responsibility and agency in criminal law, contractual obligations, and torts).

This Article attempts to diffuse this dispute by arguing that society should understand and treat AI machines, robots, agents, and algorithms (hereinafter "AI entities"[4]) as instruments of humans created and designed for the mono-purpose of being utilized by humans for their own personal benefit. A suitable analogy for AI entities must reflect its regulatory purposes. When tort law and AI collide, the appropriate regulatory goal will lead us to recognize and treat AI entities as AI judgment-proof agents, given their instrumental role in society's modern life.[5] This strongly suggests the adoption of the agency structure as the appropriate legal analogy in the AI context. In a relationship between an agent and its principal, the former is authorized to act on behalf of the latter for various purposes.[6] This Article claims the human principal is responsible for the damages caused by its AI agent given the *respondeat superior* doctrine manifested via the nature of their relationship.

The correlation between the way we view AI entities and the way we choose to analogize them—and the analogy of agency for which this Article advocates—can be viewed as creating a reinforcement circle, akin to Rawls's reflective equilibrium.[7] Reflective equilibrium manifests itself in the context of analogical reasoning in law through the process of reasoning from a particular moral judgment (i.e., particular cases) to a general one (i.e., a general theory) and back again, while both particular and general principles are adjusting and readjusting due to their mutual influence.[8] In an AI context, this process leads to the adoption of the analogy of agency given its ability to adjust itself to different circumstances and relationships in particular cases, and thus adapt the general theory to conform to these particular judgments.[9]

Human judgment-proof agents and AI judgment-proof agents (AI agents) share similar liability difficulties when damages occur. In both cases, the agent will not be held liable for the result of an accident, and his or her (or its) principal will step in as the liable party, while the agent vanishes from the legal liability picture.[10] Note that the use of the term

---

[4] The use of the phrase "AI entities" is derived from its relative neutrality. AI entities are not always robots nor are they algorithms. For more on this, see *infra* Section IV.A.

[5] Daniel Carvell et al., *Accidental Death and the Rule of Joint and Several Liability*, 43 RAND J. ECON. 51,52–54 (2012).

[6] *See infra* Part III, Section C; Part IV.

[7] JOHN RAWLS, A THEORY OF JUSTICE 48–53 (1971).

[8] Larry Alexander, *The Banality of Legal Reasoning*, 73 NOTRE DAME L. REV. 517, 518 (1998).

[9] *See* Cass R. Sunstein, *On Analogical Reasoning*, 106 HARV. L. REV. 741, 750–754 (1993) (discussing the theory of reflective equilibrium in the context of analogical reasoning).

[10] *See* RESTATEMENT (THIRD) OF AGENCY § 7.03 (AM. LAW. INST. 2006) (describing when a principal is liable to a third party harmed by an agent's conduct); *see also* Will Kenton, *Judgment Proof*, INVESTOPEDIA (June 12, 2018),

"agent" in this sense does not attribute any human-like rights or obligations to the AI entity. On the contrary, the usage of the term "agent" in the context of this Article rejects the notion that AI entities should be held liable. When it comes to identifying the liable party, the AI agent is replaced by its human principal.

The other analogies that will be presented in this Article are products, animals (domesticated and wild), slaves, electronic persons, and a grander analogy to the aviation and vaccination industries.[11] These analogies are only appropriate in the context of a full agency relationship, which is not the case when it comes to AI entities. Non-agency legal analogies are reduced to the basic problem of judgment-proof agency, where an AI entity cannot be held liable and so a human guardian, keeper, custodian, or owner must be found liable instead in order to provide a remedy.[12] In these cases, the identity of that accountable human is reduced to that of a principal (or principals) in control of the AI agent.[13] The usage of the above non-agency legal analogies in the AI context leads to the agency structure of a principal, an agent, and a victim.[14] While the precise technical definition of "principal" may vary, those in control of AI entities will still be held liable for the actions of their AI entity.[15]

Recovery is only possible via an AI agent's human principals because AI agents are effectively judgment proof. This is because these principals are the only entity the regulator can properly incentivize to prevent damages and to invest in achieving an optimal level of activity.[16] The AI agent is not transparent to the victim because the latter cannot take any preventive measures to protect herself when confronted with a harmful AI agent.[17] The principal is the one best suited to take proactive actions to prevent potential accidents and damages, to insure it, and to attain the optimal level of activity for the AI agent.[18]

---

https://www.investopedia.com/terms/j/judgment-proof.asp [https://perma.cc/34G4-GMPG] (describing what it means for a human to be judgment-proof).

[11] *See infra* Part III.

[12] *See* Sam N. Lehman-Wilzig, *Frankenstein Unbound: Towards a Legal Definition of Artificial Intelligence*, 13 FUTURES 442, 451 (1981) (explaining how the law of agency is a useful framework for thinking about liability as it relates to AI); *see also infra* Part III, Section C.

[13] *See* Lehman-Wilzig, *supra* note 12, at 451; *see also infra* Part III, Section C.

[14] *See infra* Part IV.

[15] *See infra* Part IV, Section B.

[16] *See* Harold Laski, *The Basis of Vicarious Liability*, 26 YALE L.J. 105, 112 (1916) [hereinafter Laski] ("If we allow the master to be careless of his servant's torts we lose hold upon the most valuable check in the conduct of social life.").

[17] Jack M. Balkin, *The Three Laws of Robotics in the Age of Big Data*, 78 OHIO ST. L.J. 1217, 1222 (2017) (discussing the asymmetry of information between AI and end user and arguing this creates a fiduciary relationship).

[18] *See* Laski, *supra* note 16, at 114 ("If [the principal] is compelled to bear the burden of his servant's torts even when he himself is personally without fault, it is because in a

In using AI agents as instruments, human beings in practice, advertently or inadvertently, impose risks on each other. These agents cannot be viewed as entities capable of assuming legal responsibilities, but are merely a device in the hands of the principal.[19] Building on the example made by Professor Balkin—the Rabbi is the principal behind Golem, the judgment-proof agent, and is therefore liable for any damages caused by the agent.[20]

The Article continues as follows. In Part II, the Article will provide a review of the concept of legal analogies in the AI context and the justifications for and against the application of this reasoning method.[21] Part III delves into the legal analogies offered in the AI context—property (products, domesticated and wild animals, and slaves), personhood (i.e., electronic persons), agency, and the aviation and vaccination industries—and explains how these legal analogies can all be reduced to the legal structure of agency.[22] Part IV establishes the important connection between treating AI entities as agents and the implication of this policy decision on the appropriate liability regime applicable to human principals of AI agents.[23] Part IV will also discuss, *inter alia*, the applicability of the *respondeat superior* doctrine on AI agents and the identity of the principal or principals.[24]

II. THE USE OF LEGAL ANALOGIES

"Metaphors in law are to be narrowly watched, for starting as devices to liberate thought, they end often by enslaving it."[25] Even though Judge Benjamin Cardozo wrote this statement in 1926, the notion that stands behind it still echoes in today's legal reasoning process, especially with regards to new technologies. We yearn for legal analogies, fictions, and metaphors to illuminate the unknown when it intersects with the

social distribution of profit and loss, the balance of least disturbance seems thereby best to be obtained.").

[19] Balkin, *supra* note 17, at 1230 (arguing that "[t]he owner of a fiduciary duty is not the robot. It is the company that manufactures, installs, sells, and operates the robot").

[20] *Id.* at 1222.

[21] *See infra* Part II.

[22] *See infra* Part III.

[23] *See infra* Part IV.

[24] *See* RESTATEMENT (THIRD) OF AGENCY § 2.04 (AM. LAW. INST. 2006) ("An employer is subject to liability for torts committed by employees while acting within the scope of their employment."); *id.* § 7.07(1) (describing when an employee is acting within the scope of employment); *see also infra* Part IV, Sections B–C.

[25] Berkey v. Third Ave. Ry. Co., 155 N.E. 58, 61 (N.Y. 1926); *see also* A. Michael Froomkin, *The Metaphor is Key: Cryptography, the Clipper Chip, and the Constitution*, 143 U. PENN. L. REV. 709, 860 (1995) ("Nevertheless, the observation that '[t]he power of a metaphor is that it colors and controls our subsequent thinking about its subject' is particularly relevant and powerful when the law encounters a new technology.").

law.[26] While these wise words are of particular relevance in the context of emerging and proliferating technologies, especially AI, the warning embedded in them should be taken with a grain of salt.

A. Legal Analogies and the Law

Analogies play an important role when it comes to legal reasoning.[27] Using legal analogical reasoning is based on finding a unifying normative principle. This principle validates the comparison one is trying to establish, based on similarities or differences between the new subject matter and the familiar subject matter, upon which the analogy is based.[28] While Dworkin rightly states that "analogy without theory is blind,"[29] it is also true that using an inaccurate or inappropriate theory may be myopic.[30] In order to reach a conclusion based on a proposed analogy, a well-thought-out theory and correlative principle must stand behind it. This is especially true when we discuss a technological advancement unlike anything we have encountered before.[31] It is important to note that the usage of analogy is inherently and inevitably evaluative.[32] The chosen analogy will eventually shape the desired applicable legal rules.[33]

---

[26] The words "analogy," "fiction," and "metaphor" will be used interchangeably in Part II of this Article because the referenced works do the same. However, subsequent sections will exclusively use the word "analogy," as it is the most accurate for this article's purpose. *See generally* Jacob M. Carpenter, *Persuading with Precedent: Understanding and Improving Analogies in Legal Argument*, 44 CAP. U. L. REV. 461, 471 (2016) (discussing the difference between legal analogies and legal metaphors).

[27] *See, e.g.*, LON L. FULLER, LEGAL FICTIONS (1967); PIERRE OLIVIER, LEGAL FICTIONS IN PRACTICE AND LEGAL SCIENCE (1975); LEGAL FICTIONS IN THEORY AND PRACTICE (Maksymilian Del Mar & William Twinig eds., 2015); Kenneth Campbell, *Fuller on Legal Fictions*, 2 L. & PHIL. 339 (1983); Louise Harmon, *Falling Off The Vine: Legal Fictions and the Doctrine of Substituted Judgment*, 100 YALE L.J. 1 (1990); Daniel Klerman, *Legal Fictions as Strategic Instruments* (UC Berkeley: Berkeley Program in Law and Economics, 2009); Nancy J. Knauer, *Legal Fictions and Juristic Truth*, 23 ST. THOMAS L. REV. (2010); Ben W. Palmer, *Legal Fictions and Red Room Wine: An Excursion into History*, 38 A.B.A.J. (1953); Peter J. Smith, *New Legal Fictions*, 95 GEO. L.J. 1435 (2007); Aviam Soifer, *Reviewing Legal Fictions*, 20 GA. L. REV. 871 (1986).

[28] Cass R. Sunstein, *Of Artificial Intelligence and Legal Reasoning*, 8 U. CHI. L. SCH. ROUNDTABLE 29, 32 (2001); *see also* Scott Brewer, *Exemplary Reasoning: Semantics, Pragmatics and the Rational Force of Legal Argument by Analogy*, 109 HARV. L. REV. 923 (1996).

[29] Ronald Dworkin, *In Praise of Theory*, 29 ARIZ. ST. L.J. 353, 371 (1997).

[30] Peter de Marneffe, *But Does Theory Lead to Better Legal Decisions?: Response to Ronald Dworkin's* In Praise of Theory, 29 ARIZ. ST. L.J. 427, 428 (1997).

[31] Ryan Calo, *Robots as Legal Metaphors*, 30 HARV. J.L. & TECH. 209, 215 (2016) (arguing that it is an intuitive part of the legal reasoning process to turn to legal analogies in an attempt to find the best-suited legal analogy for new phenomena in a specific legal context, particularly in the context of robots).

[32] Sunstein, *supra* note 28, at 31.

[33] *Id.*

It is hard to find an analogy that is not colored by the attitudes of the person proposing and evaluating it. The proponent is trying to generate a principle that best justifies her proposal, but there is more than one principle, in the form of a regulatory purpose, which is available per instance. It is a normative policy decision, rather than a question of similarity or dissimilarity, to support one principle over the other. This generation of competing principles for analogical reasoning is an important part of how we face new and convoluted phenomena,[34] by applying what we know and its underlying legal principle to something we have yet to master.[35]

The process of choosing the appropriate legal analogy is based upon a back-and-forth reasoning between the available analogies at hand and the regulatory purposes that they are based upon.[36] This process analyzes the nature of the problem in front of us, which is the impetus for embarking on this reasoning voyage. In an AI context, the nature of the problem is the ever-growing instances of AI entities inflicting physical injuries and non-physical damages upon society, individuals, and property, with no direct human party that can be named as liable given the unpredictable nature of these entities.[37]

After identifying the problem, we must evaluate the principles and values at stake, which will guide us in our dialectical reasoning journey between the proposed analogies.[38] An analogy on its own cannot tell us which two sets of cases are more similar (e.g. why *X* is more like *Y* than *Z*). Only a theory based on the principles and values at stake can address this task.[39] This process will lead us to an appropriate solution based on a theory consistent with the regulatory purposes we set up for ourselves.[40] It will also enable us to articulate the important principles we wish to pursue.[41] This process of reasoning is akin to Rawls's reflective equilibrium.[42] It suggests a process which is in constant movement

---

[34] *Id.* at 34 (stating that as soon as they possess evaluative capabilities, computer programs and AI will be able to do this analogical reasoning by themselves).

[35] The basic principle of using existing knowledge to analyze what will come is immensely important in the field of AI and machine learning, and it lays the foundation for its predictive purposes. It is interesting to see, therefore, how this principle is also significant in the process of cherry-picking the appropriate analogy.

[36] Sunstein, *supra* note 28, at 32–33.

[37] Jin Yoshikawa, *Sharing the Costs of Artificial Intelligence: Universal No-Fault Social Insurance for Personal Injuries*, 21 VAND. J. ENT. & TECH. L. 1155, 1162–63 (2019).

[38] Dworkin, *supra* note 29, at 355–57.

[39] *Id.* at 371–72; Larry Alexander, *Bad Beginnings,* 145 U. PA. L. REV. 57, 83 (1996).

[40] Dworkin, *supra* note 29, at 357.

[41] *Id.*

[42] *See* RAWLS, *supra* note 7, at 48; Emily Sherwin, *A Defense of Analogical Reasoning in Law*, 66 U. CHI. L. REV. 1179, 1181 (1999). An example of utilizing the reflective equilibrium process can be found in the proposal for an "analogy breaker." *See* Luke M. Milligan, *Analogy Breakers: A Reality Check on Emerging Technologies*, 80 MISS. L.J. 1319, 1320 (2011). Milligan discusses the intersection of criminal procedure and

between general principles and specific cases.[43] It uses experience and a continual loop of attempts and failures in order to investigate, reach, and eventually achieve the desired equilibrium between the general principals we as a society hold, and the specific ones that arise in idiosyncratic cases.[44]

The best analogy will not present itself problem-free. Every analogy comes with its unique deficiencies because it is an attempt to impose a relationship on two different things—it is *just* an analogy and is not identical to the subject at hand.[45] The question is, whether this deficiency undermines the applicability of the analogy as a whole, i.e., the general principle of reflective equilibrium, or creates a productive situation in the form of a manageable problem.[46] This problem will be resolved using the same back-and-forth reasoning which led us to that specific analogy in the first place.[47]

An example of the reflective equilibrium voyage can be seen with the emergence of the internet and the World Wide Web at the end of the last century. This once-new phenomenon presented, and still

---

cyberspace technology. *Id.* He offers the implementation of a doctrinal test called "analogy breaker," which allows courts to identify misleading functional analogies by "breaking" the analogy and directing their resources back to the practical implications of the new technology. *Id.* This could also be of value in a tort context to make sure the chosen analogy is substantially similar to, or different from, the technology in use, and more importantly, that it aligns with our regulatory purposes. The "analogy breaker" suggestion essentially utilizes Rawls's reflective equilibrium because it uses specific cases decided by the courts to help enforce our general understanding and values at stake relative to AI entities and their role in our society. Once an analogy is "broken" by court in a specific case, it has failed to reach the appropriate equilibrium we seek in order to reconcile our values with our practices, thus, reinforcing the back and forth reasoning process in pursuit of the best analogy.

[43] Brewer, *supra* note 28, at 927–28.

[44] *See* Robert Justin Lipkin, *Beyond Skepticism, Foundationalism and the New Fuzziness: The Role of Wide Reflective Equilibrium in Legal Theory*, 75 CORNELL L. REV. 810, 868–70 (1990); Brewer, *supra* note 28, at 927, 938–39. "Perhaps the most important of these unrecognized analogical arguments is the argument that proceeds by effecting a 'reflective equilibrium' between general norms and particular applications of those norms." *Id.* at 927–28. The method of reflective equilibrium is not the only path available when discussing the process of choosing legal analogies, but it is the most practical one in the context of this Article. Dan Hunter, *Reason is Too Large: Analogy and Precedent in Law*, 50 EMORY L.J. 1197, 1250–63 (2001) (reviewing other methods).

[45] *See* Sherwin, *supra* note 42, at 1183–84 (discussing objections to the analogical method).

[46] RAWLS, *supra* note 7, at 53 ("We should view a theory of justice as a guiding framework designed to focus our moral sensibilities and to put before our intuitive capacities *more limited and manageable questions for judgment*.") (emphasis added).

[47] Part III, Section A.1 gives an example of an analogy that failed in light of its deficiencies (AI entities as animals), and Part IV, Section B gives one that did not—AI entities as judgment-proof agents, which presents a productive problem in the form of choosing the appropriate human principal.

presents, various new legal challenges.[48] A few analogies were suggested based on existing legal rules in an attempt to identify a suitable one. First, the internet can be viewed as a conduit of information, similar to a series of tubes or a superhighway upon which information travels at high speed.[49] Second, the internet can be viewed as a novel space, usually referred to as *cyberspace*, which suggests that the internet resembles a specific geographic space.[50] Third, the internet can be viewed as a "real"—albeit not necessarily physical—space.[51] In each internet-related legal challenge, courts have decided which analogy is appropriate.[52] The court can then confirm the rule it has settled upon by testing it in light of the general principles and rationales that stand behind the analogy (i.e., testing it "up"), and test it against other specific cases (i.e., testing it "down").[53] If over time the chosen analogy survives this process, it is likely adequate for the problem as a whole. If not, a new review will be conducted, and a new analogy will be chosen.

This journey is characterized by the value of analogical reasoning as a flexible instrument which grows with the subject matter. It enables the reviewer to take into account both the general values and principles which stand as the basis of the new phenomenon, and at the same time, its application to specific cases. This methodical search will help us in

---

[48] *See* Jonathan H. Blavin & I. Glenn Cohen, *Gore, Gibson, and Goldsmith: The Evolution of Internet Metaphors in Law and Commentary*, 16 HARV. J.L. & TECH. 265 (2002) (exploring the evolution of metaphorical inferences as applied to the internet in legal commentary and judicial opinion).

[49] *Id.* at 269; *see* Clay Calvert, *Regulating Cyberspace: Metaphor, Rhetoric, Reality, and the Framing of Legal Options*, 20 HASTINGS COMM. & ENT. L.J. 541 (1998); s*ee also* Brookfield Communications, Inc. v. West Coast Entm't Corp., 174 F.3d 1036, 1064 (9th Cir. 1999) (comparing use of another's trademark in one's metatags to posting an incorrect billboard along the highway); Am. Libraries Ass'n v. Pataki, 969 F. Supp. 160, 173 (S.D.N.Y. 1997) (analogizing the internet to a highway to support interstate commerce argument). Courts later rejected the internet highway analogy. *See* Bihari v. Gross, 119 F. Supp. 2d 309, 319–21 (S.D.N.Y. 2000) (questioning the appropriateness of a highway metaphor as applied to metatags); Universal City Studios, Inc. v. Reimerdes, 111 F. Supp. 2d 294 (S.D.N.Y. 2000) (differentiating website links from road signs).

[50] Blavin & Cohen, *supra* note 48, at 275; *see* David R. Johnson & David Post, *Law and Borders—The Rise of Law in Cyberspace*, 48 STAN. L. REV. 1367, 1367 (1996) (arguing courts should take cyberspace seriously as a unique place); Reno v. ACLU, 521 U.S. 844, 851 (1997) (adopting the metaphor of cyberspace as a novel place existing outside territorial boundaries).

[51] Blavin & Cohen, *supra* note 48, at 283; LAWRENCE LESSIG, CODE 2.0 10-11 (2006); *Reno*, 521 U.S. at 890 (O'Connor J., concurring).

[52] *See* Sherwin, *supra* note 42, at 1182–83 (discussing courts' use of analogy in the context of property law).

[53] Brewer, *supra* note 28, at 962. Brewer refers to this process as first "abducting" from specific examples an "analogy warranting rule" and then testing it "up" against general principles and "down" against current or future hypothetical examples. *Id.*

establishing the adequate analogy in a manner that will guarantee we will not be blinded by its charms or enslaved by its limitations.

B. The Case Against Legal Analogies

The use of legal analogies come with a price. As Cardozo's statement suggests, using analogies can be tendentious; it can narrow our line of thought; it can inhibit the way we view new, and sometimes even old, phenomena; and it can constrain our creativity when dealing with a new and unfamiliar legal scenario.[54]

Back in 1861, Maine identified the use of legal fictions as one of three catalysts of legal change, the other two being equity and legislation. He stated that legal fictions were a "relatively primitive way of changing the law," after which the main focus was shifted to the other two catalysts.[55] In the 1930s, Fuller carried out extensive research on legal analogies, stating that "it seems exceedingly questionable whether it is ever truly convenient to employ a fiction where the judge introducing the reform can state the new rule in non-fictitious terms."[56] The widespread notion regarding legal fictions seemed to be that they should only be treated as a training or supporting instrument, which is necessary to establish the legal argument, but relevant only until a certain level of self-assurance was achieved: "one could tolerate [legal fictions], pragmatically, but only to a certain extent – the ideal was to do without them."[57]

Bentham is one of the strongest opponents of the institution of legal fictions.[58] He condemned even their mere use and considered their implementation to be a deception used by lawyers.[59] His main objection was institutional: he claimed the courts were essentially stealing power from the legislator via the application of legal analogies.[60]

Some scholars are skeptical of analogical reasoning as a substantive form of reasoning, advocating instead for a more straightforward analysis of a legal scenario with given values of stability and correct

---

[54] *See* Berkey v. Third Ave. Ry. Co., 155 N.E. 58, 61 (N.Y. 1926) ("Metaphors in law are to be narrowly watched, for starting as devices to liberate thought, they end often by enslaving it.").

[55] Maksymilian Del Mar, *Recovering Legal Fictions: An Introduction*, 9 INT'L J.L. CONTEXT 437, 437 (2013); HENRY MAINE, ANCIENT LAW 25–27 (1861).

[56] L. L. Fuller, *Legal Fictions*, 25 ILL. L. REV. 513, 524 (1931).

[57] *See* Del Mar, *supra* note 55, at 437.

[58] C.K. OGDEN, BENTHAM'S THEORY OF FICTION 141 (1932) ("fiction of use to justice? Exactly as swindling is to trade.").

[59] *See generally* Nomi Maya Stolzenberg, *Bentham's Theory of Fictions, A "Curious Double Language,"* 11 CARDOZO STUD. L. & LITERATURE 223, 227 (1999).

[60] FULLER, *supra* note 27, at 57; *see also* Sherwin, *supra* note 42, at 1183; Stolzenberg, *supra* note 59, at 229 (noting that in later years, Bentham's approach toward legal fictions changed and was more positive in light of utilitarian purposes); L. L. Fuller, *Bentham's Theory of Fiction by C.K. Ogden*, 47 HARV. L. REV. 367 (1933).

decision-making.[61] They point to the "inferiority of analogical decision-making to ordinary legal reasoning."[62] According to this approach, analogical reasoning is inferior because it depends on prior legal principles, and these prior decisions are surely morally questionable.[63]

In her article, Crootof discusses the limitations inherent in the use of analogy in the context of autonomous weapon systems, limitations which also pertain to the use of analogy in the context of AI entities.[64] According to Crootof, legal analogies may be misleading because they misrepresent important features which do not carry through to the analogy itself. The precision of the comparison is inherently limited; if it were not, there would be no need for analogies in the first place. Furthermore, legal analogies may constrain our imagination in approaching new legal situations in general, and new technological devices in particular. New technologies are particularly problematic, as they impede the proper regulation of these complex and unpredictable entities, entities that have only become more mature and integrated further into our society. The courts may fixate on a specific analogy that will not be a good representative of the technology over time.[65] The analogy limits our understanding of the subject at hand by highlighting in a selective way a specific aspect of an issue over others which are then marginalized.[66] Thus, they may be viewed as improperly result-focused, like drawing a target circle around the arrow after it has already hit the wall.[67]

C. The Inevitability of Legal Analogies

While there is a solid argument against using legal analogies and advocating for the adoption of legal standards instead, it is still not strong

---

[61] *See, e.g.*, FREDERICK SCHAUER, PLAYING BY THE RULES: A PHILOSOPHICAL EXAMINATION OF RULE-BASED DECISION-MAKING IN LAW AND IN LIFE 182–87 (1995); RICHARD A. POSNER, THE PROBLEMS OF JURISPRUDENCE 118–19 (1990).

[62] Sherwin, *supra* note 42, at 1184; Alexander, *supra* note 39, at 80–86. *Contra* Hunter, *supra* note 44, at 1241.

[63] Sherwin, *supra* note 42, at 1184–85.

[64] Rebecca Crootof, *Autonomous Weapon Systems and the Limits of Analogy*, 9 HARV. NAT'L SEC. J. 52, 79 (2018).

[65] Kyle Graham presented two examples of this problem in an article: defamation law in television and radio, and early airplane accidents. In both cases, courts tended to fixate on an old analogy, not considering the way the technology matures and diffuses into the market. These "simplistic early analogies do not always favor plaintiffs." *See* Kyle Graham, *Of Frightened Horses and Autonomous Vehicles: Tort Law and its Assimilation of Innovations*, 52 SANTA CLARA L. REV. 1241, 1256 (2012).

[66] Blavin & Cohen, *supra* note 48, at 267.

[67] *See* Calo, *supra* note 31, at 223 (explaining that metaphors and analogies that courts select for emerging technology can be outcome determinative). As evident here, it is difficult to avoid the usage of analogies when one is trying to explain an idea.

enough. For better or worse, legal analogies are inevitable.[68] Their opponents have not presented a better alternative and it would be counterintuitive to simply disregard them as part of the legal reasoning process,[69] particularly when decisions need to be made about new technologies. Legal analogies are an important part of our journey towards establishing legal standards. These standards cannot dispense with analogies entirely, but rather use them to their advantage for their future development and implementation.

In the context of autonomous weapon systems, analogies present different challenges from commercial everyday AI entities in scope, degree, and frequency of potential damages (the mere usage of autonomous weapon system entails unavoidable damage). For this reason, the way we examine the applications of legal analogies of AI entities can and should be different. The vast range of AI entities are used in diverse and distinct scenarios, which means that determining which analogy is appropriate depends on the circumstances and adapts itself over time based on the changing salient features of the technological devices via the reflective equilibrium process.[70]

The problem of hacking autonomous weapons is an important example.[71] The same considerations also apply to Internet of Things (IoT) devices, which inherently possess the ability to be hacked from near and afar, including the vast majority of AI entities connected to the internet by design.[72] The fact that they are based on algorithms, and can have their code rewritten, is an across-the-board problem which impacts the reflective equilibrium process for choosing the best-suited analogy.[73] It emphasizes that electronic persons, children, and animals make for poor analogies in the AI context, since AI entities lack genuine and unaffected independence, as will be discussed in Part III.[74]

This feature implies a lack of true autonomy and discretion for all AI entities. They can be controlled by their users, those who

---

[68] *See* DEDRE GENTER, THE STRUCTURE OF ANALOGICAL MODELS IN SCIENCE 1 (1980) (quoting Johannes Kepler: "And I cherish more than anything else the Analogies, my most trustworthy masters. They know all the secrets of Nature, and they ought to be least neglected").

[69] Del Mar, *supra* note 55, at 441.

[70] Crootof, *supra* note 64, at 56.

[71] *Id.* at 80.

[72] Lily H. Newman, *An Elaborate Hack Shows How Much Damage IoT Bugs Can Do*, WIRED (Apr. 16, 2018), https://www.wired.com/story/elaborate-hack-shows-damage-iot-bugs-can-do/ [https://perma.cc/6PCJ-TCUC].

[73] Zak Doffman, *FBI Issues 'Drive-By' Hacking Warning: This is How to Secure Your Devices,* FORBES (Dec. 6, 2019), https://www.forbes.com/sites/zakdoffman/2019/12/06/dont-get-hacked-warns-fbi-this-is-how-you-connect-smart-devices/#1fb079db360a [https://perma.cc/2HNT-KTST].

[74] *See infra* Part III; *see also* GABRIEL HALLEVY, LIABILITY FOR CRIMES INVOLVING ARTIFICIAL INTELLIGENCE SYSTEMS 28 (2015) (rejecting the animal analogy in the context of criminal liability for AI).

manufactured them, and even malicious third parties (i.e., hackers); thus, rendering analogies attributing full independence and free will unrestricted by others to these AI entities inappropriate. This, however, does not undermine the agency analogy as will be discussed shortly.

The complete and utter rejection of legal analogies is futile. Choosing this path will not benefit our legal toolkit. On the contrary, it will minimize it in a way that may create an irreversible gap between the legal realm and the non-legal realm. The mission of legal practitioners should be to minimize the damage bad analogies cause. The only way to accomplish this is by being aware of these disadvantages and considering them while deciding the suitable legal analogy in each new situation.[75]

III. THE LEGAL ANALOGIES IN THE RUNNING

This section embarks on the reflective equilibrium voyage by exploring the available AI legal analogies that are in contention. As was established in the previous section, this voyage will help us to identify the best-suited analogy for AI entities, which reconciles the general principle behind the usage of the analogy and its applicability in specific instances.

Many analogies have been suggested with regards to the place of AI in our legal and social system.[76] First, property, which includes a product (tool), a domestic pet, a wild animal or a slave; second, "electronic persons", including full-fledged adults and "quasi-persons", such as children and mentally damaged adults; third, the analogy this Article advocates for – AI entities as agents; fourth, the aviation and vaccination industries as an analogy to the AI industry as a whole.[77] Each of these suggestions present different advantages, difficulties, and consequences for choosing one analogy over the other. The use of any analogy is directly linked to the applicability of different liability regimes on AI entities and may be extremely useful in the context of AI damages if done in a methodical and coherent manner. This section will conclude with a quick remark about the notion that there is no adequate

---

[75] BARTLETT'S FAMILIAR QUOTATIONS 671 (13th ed. 1955) (quoting Samuel Butler: "analogy points in this direction, and though analogy is often misleading, it is the least misleading thing we have.").

[76] It is interesting to note that there is an entire, well-developed discipline of the reverse image of using AI technology to enhance legal reasoning and better equip the courts to adjudicate. *See, e.g.*, Bruce G. Buchanan & Thomas E. Headrick, *Some Speculation about Artificial Intelligence and Legal Reasoning*, 23 STAN. L. REV. 40 (1970); ANNE VON DER LIETH GARDNER, AN ARTIFICIAL INTELLIGENCE APPROACH TO LEGAL REASONING (1987); E.C. Lashbrooke, Jr., *Legal Reasoning and Artificial Intelligence*, 34 LOY. L. REV. 287 (1988).

[77] *See* Lehman-Wilzig, *supra* note 12, at 447–53 (detailing a quick review of these analogies).

analogy for AI entities; society should either treat AI entities the same way it treats any other injurer or as something completely new.

A. AI as Property

Society can see AI entities as their creator' or owners' property (chattels), analogized to products, animals (domesticated or wild), and according to some suggestions, even slaves.[78] This approach will lead to the application of product liability regulation, common law strict liability rules with regards to wild and domesticated animals, and masters' strict liability for damages caused by their slaves, given the fact the latter has no rights or duties of its own.

1. Products

Adopting the product analogy essentially acknowledges that AI entities, or at least the majority of them, are nothing more than toasters or fridges.[79] Much has been written about applying product liability to damages created by AI entities and most articles have concluded that given AI entities' unique features, most prominent of which is the "black-box" problem, product liability may not be enough to handle AI-inflicted injuries.[80]

---

[78] Enrique Schaerer, et al., *Robots as Animals: A Framework for Liability and Responsibility in Human-Robot Interactions*, THE 18TH IEEE INTERNATIONAL SYMPOSIUM ON ROBOT AND HUMAN INTERACTIVE COMMUNICATION (2009); Peter M. Asaro, *A Body to Kick, but Still No Soul to Damn: Legal Perspectives on Robotics*, *in* ROBOT ETHICS: THE ETHICAL AND SOCIAL IMPLICATIONS OF ROBOTICS 169, 178 (2011).

[79] Balkin, *supra* note 3, at 50; Ugo Pagallo, *Killers, Fridges, and Slaves: A Legal Journey in Robotics*, 26 AI & SOC'Y 347, 347 (2011).

[80] *See, e.g.*, Gary E. Marchant & Rachel A. Lindor, *The Coming Collision Between Autonomous Vehicles and the Liability System*, 52 SANTA CLARA L. REV. 1321, 1323 (2012) (discussing relevant liability doctrine and precedents from other technologies that may indicate how judges and juries are likely to allocate liability for autonomous vehicle crashes); F. Patrick Hubbard, *"Sophisticated Robots": Balancing Liability, Regulation, and Innovation*, 66 FLA. L. REV. 1803, 1820–29 (2014) (addressing ways the legal system may respond to address physical injury from large, mobile, and sophisticated robots that interact closely with humans); Sunghyo Kim, *Crashed Software: Assessing Product Liability for Software Defects in Automated Vehicles*, 16 DUKE L. & TECH. REV. 300, 309–17 (2018) (reviewing the challenges in applying existing product liability theories to accidents caused solely by software defects in fully automated vehicles); Bryant Walker Smith, *Automated Driving and Product Liability*, 2017 MICH. ST. L. REV. 1 (2017) (concluding that the current product liability regime, while imperfect, is probably compatible with the adoption of automated driving systems); Karni A. Chagal-Feferkorn, *Am I an Algorithm or a Product? When Products Liability Should Apply to Algorithmic Decision-Makers*, 30 STAN. L. & POL'Y REV. 61, 90–102 (2019) (presenting a classification system to identify the situations in which the application of product liability principles is warranted given specific features of the AI-based algorithm); Kenneth S. Abraham & Robert L. Rabin, *Automated Vehicles and Manufacturer Responsibility for Accidents: A New Legal Regime for a New Era*, 105 VA. L. REV. 127,

At the risk of immensely oversimplifying the science, AI is a machine or an algorithm which reaches conclusions and makes decisions without the intervention of humans.[81] Machine learning, a branch of AI, uses the initial code and data base to teach itself the "correct" or "best" decision.[82] As a result, the decision-making processes itself takes place in a virtual "black-box"[83] and is unknown to the human creator or user.[84] Neither the users nor the creators can fully understand the process and justification that stand at the basis of an AI decision-making process. If the AI entity is self-taught, society cannot know for certain who or what is responsible for its final decision, which may inflict damage, injuries, or harm to humans or property.

For these reasons, AI decisions are seen as opaque, unpredictable, and ultimately inexplicable.[85] The lack of foreseeability, AI entities' varying degrees of autonomy, and the absence of complete human control with regards to the potential behavior of AI entities leads to difficulty in establishing a legal nexus of causation between the victim and the tortfeasor as well as a difficulty in reasoning about causation in fact between the damage inflicted and the liable party.[86] This in turn

---

139 (2019); Omri Rachum-Twaig, *Whose Robot is it Anyway?: Liability of Artificial-Intelligence-Based Robots*, 2020 U. ILL. L. REV. (forthcoming 2020); Bryan Casey, *Robot Ipsa Loquitur*, 108 GEO. L. REV. 225, 271 (2019); Andrew D. Selbst, *Negligence and AI's Human Users*, 100 B.U. L. REV. (forthcoming 2020); Tim Engelhardt, *Who Pays? On Artificial Agents, Human Rights and Tort Law*, *in* RESEARCH HANDBOOK ON HUMAN RIGHTS AND DIGITAL TECHNOLOGY: GLOBAL POLITICS, LAW AND INTERNATIONAL RELATIONS 268, 277 (2019); Matthew U. Scherer, *Of Wild Beasts and Digital Analogues: The Legal Status of Autonomous Systems*, 19 NEV. L.J. 259, 280 (2018).

[81] Michael Copeland, *What's the Difference Between Artificial Intelligence, Machine Learning and Deep Learning?*, NVIDIA (July 29, 2016), blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/ [https://perma.cc/7W49-G4VY].

[82] *Id.*; Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. PA. L. REV. 633, 679 (2017). Other branches include *deep learning* and *neural network*. *See Q & A: The future of artificial intelligence*, people.eecs.berkeley.edu/~russell/temp/q-and-a.html [https://perma.cc/729E-GML3].

[83] Broadly, this term refers to "anything that has mysterious or unknown internal functions or mechanisms." *Black-Box*, MERRIAM-WEBSTER DICTIONARY, www.merriam-webster.com/dictionary/black%20box [https://perma.cc/K4S8-ZK94].

[84] Will Knight, *The Dark Secret at the Heart of AI*, MIT TECH. REV. (Apr. 11, 2017), www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/ [https://perma.cc/7DJK-KUUJ].

[85] Cliff Kuang, *Can A.I. Be Taught to Explain Itself?*, THE N.Y. TIMES MAGAZINE (Nov. 21, 2017), www.nytimes.com/2017/11/21/magazine/can-ai-be-taught-to-explain-itself.html?smid=tw-share&_r=1 [https://perma.cc/XA99-62KR].

[86] Matthew U. Scherer, *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies*, 29 HARV. J.L. & TECH. 353, 363 (2016); Jos Lehmann et al., *Causation in AI and the Law*, 12 ARTIFICIAL INTELLIGENCE & L. 279 (2004).

hampers the attribution of legal responsibility to a liable party, including the manufacturer of the AI entity within the product analogy.[87]

In order to successfully apply a product liability claim, the injured party must prove that a product has at least one of three "defect" categories—a manufacturing defect, design defect, or failure to provide adequate instructions or warnings.[88] The first refers to a situation in which the product "departs from its intended design even though all possible care was exercised in the preparation and marketing of the product."[89] With regards to AI entities, manufacturing defects are less likely because today's methods for manufacturing software are carried out with low error rates.[90] More importantly, most damages caused by AI entities cannot be traced back and proven to be the result of a specific manufacturing problem, provided that they performed according to their code and in accordance with the way they were manufactured and designed, which leads us to the second defect.[91]

A design defect will be examined according to the following standard: "the foreseeable risks of harm posed by the product could have been reduced by the adoption of a reasonable alternative design . . . and the omission of the alternative design renders the product not reasonably safe."[92] It will be extremely difficult to prove a design defect in an AI entity, given the black-box problem and the difficult question of whether an alternative safer design would have been possible, or whether the design defect would have been detected and replaced at the design stage of the AI entity.

The third defect will apply when "the foreseeable risks of harm posed by the product could have been reduced or avoided by the provision of reasonable instructions or warnings . . . and the omission of the instructions or warnings renders the product not reasonably

---

[87] Curtis E.A. Karnow, *Liability for Distributed Artificial Intelligences*, 11 BERKELEY TECH. L.J. 147, 181 (1996).

[88] RESTATEMENT (THIRD) OF TORTS: PRODUCTS LIABILITY § 2 (AM. LAW. INST. 1998).

[89] *Id.* § 2(a).

[90] *See, e.g.*, *AI Helps Manufacturers Identify Product Defects*, NVIDIA (June 22, 2017), https://news.developer.nvidia.com/ai-helps-manufacturers-identify-product-defects/ [https://perma.cc/DE44-HKBD] (explaining how AI takes pictures of an assembly line and highlights units that appear defective or anomalous).

[91] *See* John Villasenor, *Products Liability Law as a Way to Address AI Harms*, BROOKINGS (Oct. 31, 2019), https://www.brookings.edu/research/products-liability-law-as-a-way-to-address-ai-harms/ [https://perma.cc/VY7R-7LNY]; *see also* David C. Vladeck, *Machines Without Principals: Liability Rules and Artificial Intelligence*, 89 WASH. L. REV. 117, 125 (2014) (noting that courts may hesitate to assign liability to automated systems where credible alternative theories of liability exist).

[92] RESTATEMENT (THIRD) OF TORTS: PRODUCT LIABILITY § 2(b). This is called the "risk-utility test." *See* Aaron D. Twerski & James A. Henderson, Jr., *Manufacturers' Liability for Defective Product Designs: The Triumph of Risk-Utility*, 74 BROOK. L. REV. 1061, 1063 (2003) (describing the ubiquity of the risk utility formula as the standard by which to judge whether a product design is defective).

safe."[93] This duty refers to risks that are reasonably known at the time of the sale of the product. Not all AI entities' potential risks are foreseeable in the time of the sale,[94] especially if the AI entity is an open product which enables its users to modify it as they please.[95] Furthermore, complying with this duty can be achieved rather cheaply and easily by attaching a long list of instructions and possible warnings to each AI entity.[96]

Equally important to the inherent difficulty in proving these three defects when an AI entity is involved in an accident, is the fact that the features of AI entities are, in most cases and on diverse levels, different from those of simple products, which are the original focus of these review standards.

Firstly, the social connections created with humans lead us to treat AI entities differently than non-AI products.[97] To the humans who get emotionally attached to them, they are no ordinary products. This emotional connection, or at the very least a connection which is stronger than the one humanity shares with its ordinary products, bears on the appropriate AI analogy.

Secondly, AI entities' black-box problem, the relative ease in fulfilling the first and third product liability duties,[98] and the lack of ability to explain why an alternative design would have been safer than the current one, differentiates AI entities from other products. The ambiguity surrounding AI entities' capabilities and its unpredictable actions and damages, all suggest that even though in its essence an AI entity is a product that is being sold by a company to a consumer, it is not similar enough to be treated as such.

In their traditional meaning, products are still pieces of hardware, e.g., metal or plastic, which are rather mostly safe given their

---

[93] RESTATEMENT (THIRD) OF TORTS: PRODUCTS LIABILITY § 2(c).

[94] *See, e.g.*, Tom Simonite, *Google and Microsoft Warn that AI May Do Dumb Things*, WIRED (Feb. 11, 2019), www.wired.com/story/google-microsoft-warn-ai-may-do-dumb-things/?CNDID=47980536&CNDID=47980536&bxid=MjM5NjgxNTg3MDQ1S0&hasha=ced742d2d3a4866b2ea54ba0ca0d8621&hashb=837f5fc791ec721e36b7507adc59c7f41260f260&mbid=nl_021119_daily_list1_p2&source=DAILY_NEWSLETTER&utm_brand=wired&utm_mailing=WIRED%20NL%20021119%20(1)&utm_medium=email&utm_source=nl [https://perma.cc/TNJ4-NQBB] (explaining the general warning given by Google and Microsoft regarding their AI products).

[95] Ryan Calo, *Open Robotics*, 70 MD. L. REV 571, 596–97 (2011).

[96] *Id.* at 596.

[97] Kate Darling, *Extending Legal Protection to Social Robots: The Effects of Anthropomorphism, Empathy, and Violent Behavior Towards Robotic Objects*, *in* ROBOT LAW 213–16 (Ryan Calo, A. Michael Froomkin & Ian Kerr eds., 2016).

[98] Product liability requires some sort of causation which may be problematic to prove in the AI context. *See generally* Bryant Walker Smith, *Proximity-Driven Liability*, 102 GEO. L.J. 1777, 1778–79 (2014) (arguing that growing proximity could significantly expand sellers' point-of-sale and post-sale obligations toward people endangered by their automated products).

predictability. In contrast, AI entities are mostly software and may behave in an erratic and unpredictable manner.[99] Given these features, the product analogy should give way to an agent analogy, one which accounts for the basic differences between a fridge and an algorithm, and each of their social connections to humans.

Nonetheless, it should be noted that if an injured party can find and establish one of the three aforementioned defects once an AI entity inflicts damages, one should follow this claim to obtain remedy and may choose to add another claim based on the AI entities as an AI agent approach.[100]

2. Animals – Domesticated and Wild

Comparing AI entities to animals is an appealing and intuitive move, since this enables us to view the AI entity either as a domesticated pet (e.g. canine ownership),[101] or as a wild animal (e.g. elephants or tigers). The former choice suggests that the AI entity is (slightly more) predictable than the latter. This is even more appealing, given the strong feelings humans usually cultivate towards our AI companions, similar to how humans care for, and even love, non-human animal companions.[102]

With regards to wild animals, at common law, owners of animals would be held strictly liable for damage caused by their undomesticated animals if that harm is a cause of a "dangerous propensity characteristic of animals of that class, and liability does not rest on experience with particulate animals."[103] The strict liability regime is justified because the owner of the wild animal exposes others to abnormal risks in the form of a wild bear or monkey due to the unpredictable nature of the animal.[104]

---

[99] *See* Darling, *supra* note 97, at 216–17 (describing how robotic objects differ from other inanimate objects).

[100] *See infra* Part III, Section C; Part IV.

[101] Richard Kelley et al., *Liability in Robotics: An International Perspective on Robots as Animals*, 24 ADVANCED ROBOTICS 1861, 1862 (2010); *see also* Sophia H. Duffy & Jamie Patrick Hopkins, *Sit, Stay, Drive: The Future of Autonomous Car Liability*, 16 SMU SCI. & TECH. L. REV. 453, 467 (2013) (analogizing autonomous cars to domestic dogs for liability purposes). This comparison does not necessarily extend to domesticated cats. *See* Henry T. Greely, *Neuroscience, Artificial Intelligence, Crisper – and Dogs and Cats*, 51 U.C. DAVIS L. REV. 2303, 2308 (2018) (examining why domesticated cats are immune from the same kinds of liability as domesticated dogs).

[102] Darling, *supra* note 97, at 217–18.

[103] SAMIR CHOPRA & LAURENCE F. WHITE, A LEGAL THEORY FOR AUTONOMOUS ARTIFICIAL AGENTS 130 (2011).

[104] *See* George P. Fletcher, *Fairness and Utility in Tort Theory*, 85 HARV. L. REV. 537 (1972) (advocating for a strict liability regime). These could be viewed as nonreciprocal risks, according to Fletcher's approach. *Id.*

The same is true with regards to domesticated animals, under the *scienter action* at common law.[105] The owner of these animals will be held strictly liable if she has knowledge of its "vicious propensity to cause injury or damage to human beings," even if no negligence can be proven.[106] The owner will be found strictly liable if she must have known or had reason to believe "of a dangerous propensity or trait that was not characteristic of a similar animal."[107]

The resemblance between AI entities and animals is clear given their erratic behavior and unpredictability as non-human entities who, despite living side by side with us, cannot fully communicate with us. In some cases, AI entities are even designed to look and behave like our beloved animals, as in the case of Spotmini, Boston Dynamics Robot Dog, and Paro the Baby Seal Robot.[108] If it is known that the AI entity is as potentially dangerous as a wild animal and that harm is likely to occur if it escapes the owner's possession, the keeper, custodian and in most cases, owner of the animal will be liable for damages it causes if it escapes, regardless of negligence.[109] The same is true with regards to an AI entity which, like most domesticated non-human animals is not considered to be dangerous, but later becomes so. If the owner becomes aware that it is dangerous, then she will have an additional obligation, and will be held liable, if it is proved that her pet had an inclination to produce unpredictable harm. Hence, these two analogies will most likely lead to a strict liability regime when AI entities cause damages. The differentiation between them, therefore, adds no clear value to the legal analysis.

Moreover, there are two fundamental difficulties in using the analogy of non-human animals in the AI context. First, even if there was any value in distinguishing between these analogies, the classification of AI entities as domesticated or wild animals is not necessarily obvious. AI entities, such as autonomous weapons, which inherently possess a higher ability to cause damages more than other AI entities, can be easily classified as a wild animal rather than a domesticated one. However,

---

[105] *See* Mulheron, *Liability for Animals*, *in* PRINCIPLES OF TORT LAW 76, 77 (Cambridge Univ. Press, 2016); *see also* Scherer, *supra* note 80, at 282 (describing the circumstances under which a dog owner will be liable for harm caused by the dog).

[106] *See* CHOPRA & WHITE, *supra* note 103, at 130; *see also* Greely, *supra* note 101, at 2307 ("The 'vicious propensity' limitation follows from a general legal approach of giving every dog 'one free bite' before its owner, who would then be on notice of the dog's disposition, is liable for its bites.").

[107] CHOPRA & WHITE, *supra* note 103, at 130.

[108] David Z. Morris, *Boston Dynamics is Gearing Up to Produce Thousands of Robots Dogs*, FORTUNE (July 21, 2018), fortune.com/2018/07/21/boston-dynamics-spotmini-robot-dog/ [https://perma.cc/S86C-92V2]; *see also PARO Therapeutic Robot*, PARO, www.parorobots.com/ [https://perma.cc/YX6J-52ZG] (giving an overview of the PARO therapeutic robot).

[109] Mulheron, *supra* note 105, at 76.

even AI entities which have a more innocent purpose can still be highly unpredictable and it will not be easy to decide which of the two categories suits them better. A domesticated animal has gone through a process of domestication or is known to be obedient from birth. These indicators do not exist in the AI context, leading to the conclusion that humans do not possess the ability to clearly distinguish a domesticated AI entity from a wild one. All AI entities were programmed by a human programmer, were trained and guided to fulfil a certain task, and thus were domesticated in the technological sense of the word.[110] The level of dangerous activity can be an indicator to the nature of the AI entity, but the fact that AI entities are inexplicable and conduct their decision-making process in a black-box makes the classification of wild and domesticated arbitrary.[111] Furthermore, looking for a "vicious propensity" in an AI entity may be extremely problematic given the forgoing arguments.[112]

Second, unlike AI entities which were specifically created to fulfill a human-designated purpose, animals, whether they are domesticated or wild, were not.[113] They have a different purpose in life besides the purpose they serve to us as humans. Their creation, i.e., birth, happened without human interference. Their unpredictability stems from the fact they are non-human animals with needs and desires separate from our needs and desires of them.[114] This gap is not properly represented by the animal analogy and will be better explained via the agency analogy.

It seems that the animal analogy is supported by different scholars due to the strong connection humans tend to cultivate with AI entities, but overlooks the many differences between them. For this reason, the animal analogy is inappropriate when it comes to AI entities causing harms. In its essence, the keeper or owner of the animal is the one strictly liable with regard to the animal's actions, while the latter is judgment-proof. This relationship in the AI context can be reduced to an agency triangle, which will better illustrate the legal connections between the parties involved.

3. Slaves

Some scholars have called for treating AI entities as their owners' slaves,[115] i.e., using an analogy to human beings who are essentially

[110] Kelley, *supra* note 101, at 1864.
[111] *Id.*
[112] CHOPRA & WHITE, *supra* note 103, at 130.
[113] *See* Kelley, *supra* note 101, at 1868 (categorizing robots' dangerousness based on the human needs they were designed to meet).
[114] *See id.* at 1864 (comparing the unpredictability of wild animals to the greater predictability of robots).
[115] Joanna J. Bryson, *Robots Should be Slaves*, 8 NAT. LANGUAGE PROCESSING 63 (2010).

property.[116] Setting aside the problematic usage of this loaded word,[117] this analogy is based on the notion that AI entities are more intelligent than a product and more human-like than animals, in some respects, but still fall short of being free and autonomous humans, and are rather someone's property.[118]

The appeal of using slavery as an analogy for AI entities is evident from Katz's work:

> The parallels with autonomous agents are clear. The analogy (like any other analogy) is not a perfect one, but comparison may be instructive. Like a slave, an autonomous agent has no right or duties itself. Like a slave, AI is capable of making decisions which will affect the rights (and, in later law, the liabilities) of its master. By facilitating commercial transactions, autonomous agents can increase market efficiency. Like a slave, an autonomous agent is capable of doing harm.[119]

To further advocate for this analogy, Pagallo presents the example of "peculium" from Ancient Rome in the contractual context.[120] Given the important role slaves had in Ancient Rome's trade and commerce, the peculium is a mechanism provided by Roman law which enables a business to be run by slaves similar to a limited company, although the business is still considered to be the property of the master.[121] Therefore, Pagallo continues, society can apply this mechanism to modern transactions mediated by AI entities. The attractiveness of using the peculium mechanism stems from the fact that it shows "a sound way to forestall any legislation that might prevent the use of robots due to their risks and consequent excessive burden on the owners . . . of robots."[122]

---

[116] Pagallo, *supra* note 79, at 351 ("[W]e may deem robots as autonomous agents that are however merely 'things.'").

[117] *See, e.g.*, Joan M. Herbers, *Watch Your Language! Racially Loaded Metaphors in Scientific Research*, 57 BIOSCIENCE 104, 104 (2007); Ron Eglash, *Broken Metaphor: The Master-Slave Analogy in Technical Literature*, 48 TECH. & CULTURAL 360, 369 (2007) ("Surely, between our cultural resources and our desire for technical accuracy, we can do better than 'master' and 'slave'"). In the American context, slavery analogies are considered taboo. This is not the case in European Civil law, which is entrenched in Roman law, where slavery has a different historical, cultural and social context and meaning. *See generally* W. W. BUCKLAND, THE ROMAN LAW OF SLAVERY: THE CONDITION OF THE SLAVE IN PRIVATE LAW FROM AUGUSTUS TO JUSTINIAN (1908); ALAN WATSON, ROMAN SLAVE LAW (1987).

[118] Pagallo, *supra* note 79, at 351.

[119] Andrew Katz, *Intelligent Agents and Internet Commerce in Ancient Rome*, SOC'Y FOR COMPUTERS & L. (Oct. 15, 2008), https://www.scl.org/articles/1095-intelligent-agents-and-internet-commerce-in-ancient-rome [https://perma.cc/BE9U-4AX9].

[120] Pagallo, *supra* note 79, at 351.

[121] *See* BUCKLAND, *supra* note 117, at 131 (describing the nature of the peculium).

[122] Pagallo, *supra* note 79, at 352.

It essentially allows limited liability for the owner of robots and a business warranty for those who have been injured by the AI entity.[123]

Although the theoretical framework of the peculium is intriguing, it fails to add value over that of legal agency, including instances of corporate agency, which stand at the center of the peculium mechanism. The AI entity is the representative of the company, its agent, and thus it limits the liability of the slave itself and deflects the liability to the principal (even if it is in times limited). The slavery analogy does not contribute to the legal analysis over and above that of using the analogy of agency.

In her article, Bryson unequivocally calls for the adoption of the legal analogy of AI entities as slaves.[124] She claims that "robots should be built, marketed and considered legally as slaves, not companion peers."[125] Furthermore, in her opinion, humans fully own robots and determine their goals.[126] Her article presents four fundamental claims: "1) Having servants is good and useful, provided no one is dehumanized. 2) A robot can be a servant without being a person. 3) It is right and natural for people to own robots. 4) It would be wrong to let people think that their robots are persons."[127]

Delving into each of these claims exceeds the scope of this Article; however, it is important to note that the analogy of a servant is not identical to that of a slave, as Bryson seems to suggest. These four claims do not necessarily lead to the notion that slaves are the most suitable analogy, but rather that of servants which are agents of their principals, even if the agent itself is not a person. Robots can be thought of as an extension of us, similar to an agent, not a slave.[128] The fact that humans dictate an AI entity's motivation, design and decision-making system, and that "all their goals are derived from us" does not necessarily lead to the notion of slavery, unless humans wish to emphasize the property aspect of AI entities at the expanse of their functionality.[129]

Nonetheless, the property aspect makes no difference to the final liability regime between the analogy of slaves and agents; whereas the former entails personal, historical, and cultural reasons to object,[130] the latter acknowledges the features Bryson identifies in her article without

---

[123] *Id.* (stating there is no clear suggestion to a liability regime in the extra-contractual aspect).

[124] Bryson, *supra* note 115, at 63.

[125] *Id.*

[126] *Id.* ("The potential of robotics should be understood as the potential to extend our own abilities and to address our own goals.").

[127] *Id.* at 65.

[128] *Id.* at 71.

[129] *Id.* at 73.

[130] *See generally supra* discussion in note 117 (describing how this occurs mostly in the American context).

diminishing AI entities to nothing but property.[131] This Article rejected this statement in its previous discussion about AI as products or animals, given their unique characteristics, which attributes to AI entities something more than being mere property in light of their ability to evolve and improve over time.[132]

The slave analogy tries to manifest both the property and the function aspect of the AI entity, but ends up alienating itself in the eyes of the public. Legal analogies are grounded on our common understanding of a new phenomenon and are not meant only for the law community, but for everyone who encounters that phenomenon. Choosing an analogy that is rejected by the public given its historical context, is not optimal especially if the same regulatory goal could be achieved by utilizing agency law, a far less controversial path. Moreover, there is merit in treating AI entities as something more than just property. This allows for flexibility in light of fast advancements in AI technology and implement those advances into the application of the chosen analogy. The slavery analogy does not offer this benefit because of its rigid property centralized context.

After reviewing the property side of the spectrum, this Article turns to the opposing contender situated on the other side—treating AI entities as electronic persons with independent rights and obligations.

B. AI as Electronic Persons

1. Quasi-Persons

AI entities could be considered as "quasi-persons" much like children, and as some may even say, unborn human fetuses and severely brain damaged and comatose individuals.[133] Children are not entitled to the same rights of personhood as adults and they have limited legal status, for example, children cannot sign contracts, drink, or vote.[134] Children receive more protections than adults in some cases, and their responsibility for damages may be diminished given certain circumstances.[135] This comparison seems to be based on the notion that both children and AI entities are only able to control their actions up to a certain degree. The rationale that stands at the basis of this analogy is that if society is willing to diminish the responsibility of children based on characteristics they allegedly share with AI entities, society should

---

[131] Bryson, *supra* note 115, at 63.

[132] *See* Darling, *supra* note 97, at 218–20 (describing humans' emotional reactions and attachments to robots with anthropomorphic qualities).

[133] Asaro, *supra* note 78, at 179; Lehman-Wilzig, *supra* note 12, at 450.

[134] Scherer, *supra* note 80, at 283.

[135] S*ee, e.g.*, GIDEON YAFFE, THE AGE OF CULPABILITY: CHILDREN AND THE NATURE OF CRIMINAL RESPONSIBILITY 66–96 (2018) (providing an example of damages in the criminal context).

extend the same courtesy and apply the same logic to AI entities, whose actions cannot always be rationalized or justified.

When a child causes damages independently—with no help or adult intervention—the focus shifts from strict liability to negligent supervision by the child's guardians.[136] This obligation acknowledges the duty of care a parent has to prevent harm, even though there is no general rule which creates an obligation to prevent damages caused by a third party (i.e., the child).[137] The child analogy tries to create a connection between the duty placed on parents "to take reasonable care to control minor children or as to prevent them from intentionally harming others or from creating an unreasonable risk of bodily harm to them,"[138] to that of an owner, operator, designer, trainer or programmer of an AI entity.[139] Similar to the case of domesticated animals, parents are only obligated by this duty if it can be shown the minor has a "propensity" to cause a specific type of damage and the parents were aware of that propensity and failed to take action to prevent the destructive behavior.[140]

This analogy is fundamentally identical to that of domesticated animals at common law. Many AI entities resemble toddlers and small children in their external appearance and in their behavior and interactions with those who surround them.[141] However, unlike pets, treating AI entities as children attributes a new human aspect to them. It suggests, that just like toddlers, AI entities will learn, develop, and grow to become a part of adult society as fully capable and independent entities. Society should sanction any mischief they might cause in a pedagogical-rehabilitative way in order to ensure they will grow up to be contributing members of society and that their guardians will make sure that will come to pass. But that is not the case when it comes to AI entities. The rationale standing behind punishing parents for negligent supervision is not similar to that of sanctioning the parallel party in the AI context. This also begs the question of who should be viewed as the appropriate supervisor (i.e., parent or guardian) in the AI context. In the child analogy, the answer is mostly clear but finding the parallel

---

[136] Greely, *supra* note 101, at 2317.

[137] CHOPRA & WHITE, *supra* note 103, at 132.

[138] *Id.* at 133; RESTATEMENT (SECOND) OF TORTS § 316 (AM. LAW INST. 1965).

[139] *See* Shlomit Yanisky-Ravid, *Generating Rembrandt: Artificial Intelligence, Copyright, and Accountability in the 3A Era – The Human-Like Authors are Already Here – A New Model*, 2017 MICH. ST. L. REV. 659, 691 (2017) (discussing the multi-player model in a copyright context).

[140] CHOPRA & WHITE, *supra* note 103, at 133–34. *See also* PAGALLO, *supra* note 3, at 125. The latter also elaborates on the Italian Civil Code and the civil law approach to the field of extra-contractual obligations. *Id.* at 126–30.

[141] *See* Rachel England, *Adorable Home Robot Kuri Is Being Discontinued*, ENGADGET (July 25, 2018), https://www.engadget.com/2018/07/25/adorable-home-robot-kuri-is-being-discontinued/ [https://perma.cc/NP89-XTTC] (showing a picture of "Kuri," which was intended to enter the home as a family member).

entity in the AI realm is a difficult question, one that cannot be resolved using the narrow analogy of AI-as-child.

Like animals, children were never meant to fulfill fully mature humans' commands and requests, even if they do resemble children in their ability to learn from their mistakes to improve their future behavior via a trial and error process (i.e., reinforcement learning). The gaps in the means, purpose, and agenda are too great to be subsumed within a single category. Again, in the AI context, using the analogy of agency better represents the relationship of the relevant parties. The "AI child" is a judgment-proof agent of its custodian. Therefore, the agency analogy is better equipped to regulate AI accidents than that of AI-as-child.

2. AI Personhood

The most liberal analogy for AI entities is that of "electronic persons" with rights and obligations (i.e., treating AI entities as equals to human adults).[142] This approach openly embraces the homunculus idea, according to which the AI entity is a legal being, separate from its creators owners, or dispatchers, and is by itself responsible for its behavior, actions, and consequences.[143]

The category of "electronic persons" grants an AI entity personhood as a member of society.[144] This personhood may include obtaining citizenship (as in the case of the robot Sophia, which received Saudi Arabia citizenship in October 2017), paying taxes, exercising First Amendment rights, exercising intellectual property rights, and the ability to express consent.[145]

---

[142] Leon E. Wein, *The Responsibility of Intelligent Artifacts: Toward an Automation Jurisprudence*, 6 HARV. J.L. & TECH. 103, 140 (1992).

[143] Balkin, *supra* note 3, at 58–59.

[144] *See Report of the Committee on Legal Affairs with Recommendations to the Commission on Civil Law Rules on Robotics*, at 18 (Jan. 1, 2017), http://www.europarl.europa.eu/doceo/document/A-8-2017-0005_EN.pdf [https://perma.cc/EE2N-YCBE] [hereinafter *EU Report*].

[145] *See* Kristen Korosec, *Saudi Arabia's Newest Citizen is a Robot*, FORTUNE (Oct. 26, 2017), fortune.com/2017/10/26/robot-citizen-sophia-saudi-arabia/ [https://perma.cc/4DH5-HLB7]; Robert David Hart, *Saudi Arabia's Robot Citizen is Eroding Human Rights*, QUARTZ (Feb. 14, 2018), qz.com/1205017/saudi-arabias-robot-citizen-is-eroding-human-rights/ [https://perma.cc/D2SU-SG4G] (giving a critical aspect of granting citizenship to robots); Ryan Abbott & Bret N. Bogenschneider, *Should Robots Pay Taxes? Tax Policy in the Age of Automation*, 12 HARV. L. & POL'Y REV. 145, 168 (2018) (describing tax policy options for an automation tax); Ignatius Michael Ingles, *Regulating Religious Robots: Free Exercise and RFRA in the Time of Superintelligent Artificial Intelligence*, 105 GEO. L. REV. 507, 519–20 (2017) (outlining arguments for and against giving robots free exercise of religion protection); Russ Pearlman, *Recognizing Artificial Intelligence (AI) as Authors and Inventors Under U.S. Intellectual Property Law*, 24 RICH. J.L. & TECH. 1, 3–4 (2018) (advocating for intellectual property rights for AI); Lily Frank & Sven Nyholm, *Robot Sex and Consent:*

In a liability context, this legal analogy may lead to sanctioning the AI entity itself. This result seems to resemble a piece of science fiction that is impractical and lacking legal justification, and which stands at the heart of the *ex-post* and *ex-ante* systems (e.g., deterrence and fairness).[146] Nonetheless, this approach has gained much traction in light of the EU Parliament report published in 2016, which suggested treating AI entities like electronic persons as a possible future solution to AI liability.[147] Many read into this suggestion the possibility of granting robots their own "personhood" as separate legal beings with obligations and, as a result, perhaps even deserving rights.[148] This led to scholarly turmoil and protest. More than 150 experts in robotics, artificial intelligence, law, medical science, and ethics warned against Europe granting robots rights.[149] In an open letter to the European Commission, these experts claimed the proposal appears to be influenced more by science fiction than real world experience.[150]

However, it is important to note that the EU report does not display favoritism towards one solution over another; it merely tries to provide a general futuristic picture of what instruments of legislation the EU can consider in attempting to regulate robots in a liability context.[151] Furthermore, the report states nothing about granting AI rights but rather only talks about their obligations.[152] The problem with this argument, like that presented in the context of animals, is that the existence of obligations usually leads to the parallel existence of rights.[153]

Is Consent to Sex between a Robot and a Human Conceivable, Possible, and Desirable?, 25 ARTIFICIAL INTELLIGENCE & L. 305 (2017).

[146] *See* GUIDO CALABRESI, THE COSTS OF ACCIDENTS – A LEGAL AND ECONOMIC ANALYSIS 35–95 (1970) (explaining the goals and sub-goals of the tort system); HALLEVY, *supra* note 74, at 185–229 (explaining the sanctioning of AI entities in the criminal context).

[147] *See EU Report*, *supra* note 144, at 18 ("Creating a specific legal status for robots in the long run, so that at least the most sophisticated autonomous robots could be established as having the status of electronic persons responsible for making good any damage they may cause, and possibly applying electronic personality to cases where robots make autonomous decisions or otherwise interact with third parties independently.").

[148] *Id.*

[149] Ivana Kottasová, *Experts Warn Europe: Don't Grant Robots Rights*, CNN (Apr. 12, 2018), money.cnn.com/2018/04/12/technology/robots-rights-experts-warn-europe/index.html [https://perma.cc/NVS7-PVM8]; Janosch Delcker, *Europe Divided over Robots 'Personhood'*, POLITICO (Apr. 11, 2018), www.politico.eu/article/europe-divided-over-robot-ai-artificial-intelligence-personhood/ [https://perma.cc/TL5H-3DQ7].

[150] ROBOTICS–OPEN LETTER, www.robotics-openletter.eu/ [https://perma.cc/5CEG-UDNE].

[151] *See EU Report, supra* note 144, at 17.

[152] *Id.* at 36.

[153] In the animal context, some argue for granting non-human animals rights, but if we can sue humans for abusing or killing animals, then theoretically the opposite is also

It is hard to impose obligations on the one hand without granting rights on the other. This may lead to a slippery slope where imposing obligations on AI entities will inevitably lead to also assigning them rights.[154] For example, if an AI entity is found liable for the damage it causes, it seems only reasonable that a human can be found liable if she causes damages to the AI entity itself.[155] In the AI context, society is trying to protect humans from potential damages they may be exposed to by AI entities, and in the process, society may unintentionally reward the latter rights that it did not want them to have.

In April 2018, the EU Commission published an outline of an artificial intelligence strategy which does not mention the European Parliament call to grant personhood for AI.[156] This does not necessarily entail the complete rejection of the notion, but rather the rejection of its immediate adoption, as can be expected given the current technological status of AI entities, and the gap between their capabilities and human capabilities.[157]

On the other hand, some scholars argue that awarding AI entities with personhood will be unavoidable in the future due to the fast advancement of AI technology.[158] Chopra and White argue that "the conditions for each kind of legal personality could, in principle, be met by artificial intelligence agents in the right circumstances. We suggest that objections to such a status for them are based on a combination of human chauvinism and a misunderstanding of the notion of a legal person."[159] Chopra and White suggest this is a pragmatic question rather

---

true; we could sue a non-human animal attacker that has injured a human. Without knowing what rights we owe animals, it is difficult to know what rights we would owe robots. Nathan Heller, *If Animals Have Rights, Should Robots?*, THE NEW YORKER (Nov. 28, 2016), www.newyorker.com/magazine/2016/11/28/if-animals-have-rights-should-robots [https://perma.cc/68HR-YSR2].

[154] *See id.*

[155] *See* Todd Leopold, *HitchBot, the Hitchhiking Robot, Gets Beheaded in Philadelphia*, CNN (Aug. 4, 2015), www.cnn.com/2015/08/03/us/hitchbot-robot-beheaded-philadelphia-feat/index.html [https://perma.cc/P4X4-CEDP] (discussing a case involving human vandalism of a robot).

[156] European Commission Press Release IP/18/3362, Artificial Intelligence: Commission Outlines a European Approach to Boost Investment and Set Ethical Guidelines (Apr. 25, 2018), europa.eu/rapid/press-release_IP-18-3362_en.htm [https://perma.cc/3LGC-ME8F].

[157] Thomas Burri, *The EU is Right to Refuse Legal Personality for Artificial Intelligence*, EURACTIV (May 31, 2018), www.euractiv.com/section/digital/opinion/the-eu-is-right-to-refuse-legal-personality-for-artificial-intelligence/ [https://perma.cc/XH9S-VLRB] (stating the refusal to immediately adopt the idea is correct because: (1) the EU cannot establish what "personhood" means (it is up to the individual member states); and (2) giving personhood to AI would lead to unintended consequences).

[158] Amanda Wurah, *We Hold These Truths to Be Self-Evident, that All Robots are Created Equal*, 22 J. FUTURES STUD. 61, 64 (2017); CHOPRA & WHITE, *supra* note 103, at 191.

[159] CHOPRA & WHITE, *supra* note 103, at 27.

than a logical one: "considering artificial agents as legal persons is, by and large, a matter of decision rather than discovery, for the best argument for denying or granting artificial agents' legal personality will be pragmatic rather than conceptual."[160] They claim that the law might require the future recognition of AI as legal persons in light of their functionality and social role. Despite this, they still reject the notion of AI personhood due to a pragmatic approach, but state they do not object to robot's personhood in the future when the conditions are ripe.[161] Chopra and White conclude by asserting that the only cause that may deprive the acknowledgment of robots as legal persons is "a chauvinistic preservation of a special statues for biological creatures' uniqueness and the singularity of the human mind and moral sense in a naturalistic world order."[162] In essence, they claim that if society will not grant rights to robots in the future, it will be due to our condescending nature and the belief that the human race is superior to all other species.

Pagallo also discusses this topic by presenting Solum's approach from 1992 and his attempt to refute the three main arguments against granting AI legal personhood: first, AI entities are not humans; second, the "missing-something" argument; third, AI entities ought to be property.[163] Pagallo presents the arguments made by "the front of robotic liberation."[164] First, legal personhood "might prevent the ethical aberration of robots being treated as mere slaves."[165] Second, granting legal personhood to robots "would provide a more coherent picture of today's legal framework."[166] Pagallo eventually rejects the notion of legal personhood for robots: "for the foreseeable future, it is thus likely that the independent personhood of robots will not be on the legal agenda."[167] Following a logic similar to that of Chopra and White, abovementioned, he determines that treating robots as legal persons in the fields of immunity, strict liability, and unjust damages raises the discussion to a level of science fiction and is therefore inadequate to be applied on AI entities in today's legal environment.[168]

Treating AI entities as persons and granting them AI personhood would be getting ahead of ourselves in light of today's technological advancements. This does not mean it will be irrelevant in the future, but rather that, as of now, treating AI entities as humans will be grossly

---

[160] *Id.* at 154.

[161] *Id.* at 159. "Only the first kind [dependent legal personality] is likely to be accorded to artificial agents unless or until they attain a very high degree of autonomy." *Id.*

[162] *Id.* at 191.

[163] PAGALLO, *supra* note 3, at 158–60 (citing Lawrence B. Solum, *Legal Personhood for Artificial Intelligences*, 70 N.C. L. REV. 1231, 1258–76 (1992)).

[164] *Id.* at 155.

[165] *Id.* at 158–60.

[166] *Id.*

[167] *Id.* at 165.

[168] *Id.* at 164–65.

misguided given their dependency on an owner, operator, designer, trainer, or programmer in the process of their decision making. There might be a future where AI entities can and should be treated as equals to humans, but that future is still far off.[169] Trying to impose it given today's state of technological progress would be inaccurate and premature.[170]

C. AI as Agents – Agency Law and Corporations

This Article advocates employing an analogy that treats AI entities as agents, as if they were "servants" of their owners, operators, designers, trainers, or programmers.[171] The agency analogy refers to AI entities as agents which are monitored, guided, and receive directions and tasks from a human principal who has control over the purposes the AI entity has to accomplish.

This analogy includes corporations, which are considered separate legal beings and can be sanctioned even though they are not human. Agency law essentially allows corporations to act.[172] The corporation functions as a legal entity for all intents and purposes; it can sign contracts, sue and be sued, and buy property, among other things.[173] The corporation acts via its board of directors (the agents), who were authorized by the company itself via its shareholders (the principals).[174] The company will be held liable if its agents cause damages.[175] In case the company is insolvent (as is the case for AI entities, which are always insolvent), a sanction will be imposed on the company shareholders behind its corporate veil, not on the company itself or the agents.[176] This is a good example of the way in which agency relationship could be constructed even though one of the parties is not a human entity. In

---

[169] *See* Scherer, *supra* note 80, at 262–63 (noting a similar conclusion to that of Pagallo).

[170] *See generally* Mark Coeckelbergh, *Robot Rights? Towards a Social-Relation Justification of Moral Consideration*, 12 ETHICS & INFO. TECH. 209 (2010); F. Patrick Hubbard, "*Do Androids Dream?": Personhood and Intelligent Artifacts*, 83 TEMP. L. REV. 405 (2011); Phil McNally & Sohail Inayatullah, *The Rights of Robots: Technology, Culture, and Law in the 21st Century*, 20 FUTURES 119 (1988); Herman T. Tavani, *Can Social Robots Qualify for Moral Consideration? Reframing the Question about Robot Rights*, 9 INFO. 73 (2018); David J. Gunkel, *The Other Question: Can and Should Robots Have Rights?*, 20 ETHICS & INFO. TECH. 87 (2018); Robert van den Hoven van Genderen, *Do We Need New Legal Personhood in the Age of Robots and AI?*, *in* ROBOTICS, AI AND THE FUTURE OF LAW 15 (Marcelo Corrales et al. eds., 2018).

[171] Lehman-Wilzig, *supra* note 12, at 451. *But see* Rachum-Twaig, *supra* note 80 (stating new laws should be considered to address liability related to AI).

[172] RESTATEMENT (THIRD) OF AGENCY § 1.01 (AM. LAW INST. 2006).

[173] *Id.*

[174] *Id.*

[175] RESTATEMENT (THIRD) OF AGENCY § 7.03 (AM. LAW INST. 2006).

[176] *See* David K. Millon, *Piercing the Corporate Veil, Financial Responsibility, and the Limits of the Limited Liability*, 56 EMORY L.J. 1305, 1307 (2007).

both cases—corporate principals and AI agents—the human which stands behind the scenes is the true puppeteer.[177]

Agency law and the legal doctrine of vicarious liability (*respondeat superior*) go hand-in-hand in the tort context when damages occur.[178] A legal three-party relationship is formed when an agent—in our case an AI entity—carries out an order from its superior which inflicts damage on a third party or a third party's property.[179]

It is important to note that the *Restatement (Third) of Agency Law* states in its terminology clause that computer programs cannot be considered as agents.

> [A] computer program is not capable of acting as a principal or an agent as defined by the common law. At present, computer programs are instrumentalities of the persons who use them. If a program malfunctions, even in ways unanticipated by its designer or user, the legal consequences for the person who uses it are no different than the consequences stemming from the malfunction of any other type of instrumentality. That a program may malfunction does not create capacity to act as a principal or an agent.[180]

One conclusion that can be derived from this comment is that current legislation binds software to be nothing more than an instrument in a way that its degree of "intelligence" is not relevant.[181] However, the wording of this comment articulates the notion that stands behind our justification to treat AI entities as agents—it is indeed an instrument in the hands of its human controller, but it is not only that.[182] If society accepts this comment at face value, the discussion surrounding AI's inflicting damages is superfluous, because computer software should also be treated as a product. This is not necessarily true. The comment written in 2006 states that "at present" computer programs are nothing more than an instrument, but AI advancements may disprove this assumption in light of new factual and technological advancements, which create a new "at present" assumption of today's technology and that of the years to come.[183]

---

[177] PAUL L. DAVIS, GOWER'S PRINCIPLES OF MODERN COMPANY LAW 201–33 (6th ed., 1997). *See also* MARC MOORE & MARTIN PETRIN, CORPORATE GOVERNANCE: LAW, REGULATION AND THEORY 35–37, 230–34, 268–70 (2017) (discussing the "agency costs" in corporations); Scherer, *supra* note 80, at 264 (discussing AI personhood through a corporation analogy).

[178] *See* discussion *infra* Part IV, Section B.4.

[179] DAVIS, *supra* note 177, at 176. *See also* P. S. ATIYAH, VICARIOUS LIABILITY IN THE LAW OF TORTS (1967); CHOPRA & WHITE, *supra* note 103, at 128 (providing additional discussion regarding *respondeat superior*).

[180] RESTATEMENT (THIRD) OF AGENCY § 1.04 cmt. e (AM. LAW INST. 2006).

[181] Shawn Bayern, *Artificial Intelligence and Private Law*, *in* RESEARCH HANDBOOK ON THE LAW OF ARTIFICIAL INTELLIGENCE 144, 150 (Woodrow Barfield & Ugo Pagallo eds., 2018).

[182] Scherer, *supra* note 80, at 262 (noting AI can also do things without human control).

[183] *Id.* at 262–63.

The comment states:

> [i]f a program malfunctions, even in ways unanticipated by its designer or user, the legal consequences for the person who uses it are no different than the consequences stemming from the malfunction of any other type of instrumentality.[184]

If society agrees to this presumption, then there is nothing particularly special or novel about the way the law should treat AI. Nonetheless, other factors suggest otherwise and have demonstrated that this is not the case. Among them are the vigorous academic and technological discussions surrounding AI, AI's unique features and roles in various industries, and AI's ability to reach decisions in a different way than popular computer software from more than a decade ago.[185] If the agency analogy is accepted, this legislation should be amended accordingly and not prevent the treatment of AI entities as AI agents.[186]

When it comes to autonomous technology in general, and AI-specific applications, Chopra and White make a strong case for autonomous artificial agents.[187] Calo also alludes to robots as agents when he presents the example of using the robot metaphor to describe an entity as an extension of a person.[188] Treating such an entity as someone's robot essentially views it as an agent—the long hand of the principal. This analogy is also appropriate even if the robots are not programmable machines and possess some degree of independent "will," and in so doing, they resemble people more than some may think or want to believe.

---

[184] RESTATEMENT (THIRD) OF AGENCY § 1.04 cmt. e (AM. LAW INST. 2006).

[185] Scherer, *supra* note 80, at 263. *See also* Salil K. Mehra, *Antitrust and the Robo-Seller: Competition in the Time of Algorithms*, 100 MINN. L. REV. 1323, 1366–67 (2016) (discussing this comment in an antitrust context); David Marc Rothenberg, *Can Siri 10.0 Buy Your Home: The Legal and Policy Based Implications of Artificial Intelligent Robots Owning Real Property*, 11 WASH. J.L. TECH. & ARTS 439, 449–51 (2016) (arguing that in practice, robots are already acting as agents despite the wording of the Restatement).

[186] Scherer, *supra* note 80, at 286 ("[T]he fact that A.I. systems are not legal 'persons' would present no barrier to treating them as agents of the persons or entities who created or deployed them.").

[187] CHOPRA & WHITE, *supra* note 103, at 154. *See also* PAGALLO, *supra* note 3, at 166. Chopra and White advocate for agency, stating the hardest problem in assigning liability to artificial agents is determining which case law to use. CHOPRA & WHITE, *supra* note 103, at 119. They compare liability for artificial agents to that of animals, children, prisoners (or slaves), and extremely hazardous activities. *Id.* at 120. *See also* Pagallo, *supra* note 79, at 353; Scherer, *supra* note 80, at 290 (stating treating AI entities as agents "strike a balance
between ensuring that victims receive compensation and protecting designers and sellers of A.I. systems from liability for harm that results from transformations (as opposed to mere implementations) of their technologies.").

[188] Calo, *supra* note 31, at 221.

AI entities present challenges in the form of enhanced capabilities and unpredictability, thus outgrowing the analogy of being mindless. Calo offers some insight to the way the robot itself is used as a legal metaphor by courts and officials,[189] usually utilizing features attributed to robots and robot-like behavior to human beings when they want to describe the "shorthand of a person without will," which is completely controlled by another.[190] This notion of robotic behavior has become inaccurate as AI entities' abilities have increased over recent years.[191] The back and forth reasoning between the cases adjudicated in court and the general principles we hold with regards to AI entities has shifted with the advancement of AI technology from the analogy of mere products to something more, without fully understanding what exactly that addition is and how it can be manifested.

The AI-as-agent scheme seems preferable compared to other suggested legal analogies, in part because the latter eventually converge on the agency model.[192] Furthermore, it offers greater adaptability and

---

[189] *Id.* at 216 (stating judges may "invoke emerging technology rhetorically in order to motivate or justify decisions about people").

[190] *Id.* at 223.

[191] *Id.* at 224.

[192] *See supra* Part III (covering these analogies and their convergence); *see also* Ignacio N. Cofone, *Servers and Waiters: What Matters in the Law of A.I.*, 21 STAN. TECH. L. REV. 167 (2018) (discussing a new way to treat AI entities by using legal analogies based on three AI traits derived from Calo and Balkin: embodiment, social valence, and emergence). *See also* Ryan Calo, *Robotics and the Lessons of Cyberlaw*, 103 CALIF. L. REV. 513, 545 (2015); Balkin, *supra* note 3, at 57. Emergence is divided into two aspects—unpredictability and agency. Cofone, *supra* note 192, at 183. The first aspect focuses on the human perspective and examines humans' ability to foresee (direct liability) or control (vicarious liability) AI entities to determine humans' liability when an AI entity inflicts damages. The second aspect uses the term "agency" to refer to the internal perspective of the robot itself to determine the extent to which AI entities "could respond to legal incentives directly." *Id.* at 185. In other words, this perspective asks whether (and to what extent) an AI entity operates autonomously and independently. Conversely, this Article's discussion of agency refers to the relationship between a principal and an agent. In the context of Cofone's article, "agency" is used to talk about the autonomous ingredient an AI entity will possess in the future, because "AI agents with agency are so far inexistent." *Id.* Cofone's proposed framework would select the appropriate analogy for every AI accident case. *Id.* at 174. The analogies Cofone suggests represent a continuum—from an adult human, to a child, to a domesticated animal, to a wild animal, and finally, to a corporation or a tool. *Id.*
However, the spectrum between adult humans—for which there is not currently a robotic equal—and tools and products could be represented by an agent-principal relationship for two reasons. First, the levels of embodiment, emergence, and social valance (very high, high, low, and very low, respectively) can be manipulated to achieve a pre-determined analogy. Deciding the degree of a feature in an AI entity may be complicated and experts from differing fields (e.g., science and law) may interpret each feature-level in a different way relative to the same AI entity and circumstances. *Id.* at 176. Second, until agency as described by Cofone materializes, analogizing AI entities to adult humans—or even children—is premature. *Id.* at 196.

flexibility relative to the other contenders.[193] One of its major features is its ability to act according to circumstances and function in a context-specific way.[194] The more unpredictable an AI entity becomes, the more it is akin to a human agent, but it does not amount to a fully-fledged human.[195] AI entities are proactive, have a varying degree of social ability to interact with humans, and are independent from humans to a certain extent, that is, there is a low level of human intervention in its decision-making process.[196] Therefore, the analogy to a judgment-proof agent provides a model that fundamentally presents the legal relationships of the other available analogies in an accurate way. It also offers a wide range of functions and context-based scenarios that are unparalleled to other instances of legal analogies, such as pets and children.[197]

Human agents exist in our world in many frameworks, as drivers, delivery men and women, employees, trustees, and more.[198] This existing body of agents can be extremely helpful when we shift our focus to AI entities carrying out activities that may cause damages. Furthermore, many AI entities are created to eliminate the need for human agents as an AI machine is offered to complete the job instead.[199] This also strengthens the appropriateness of using the analogy of agents in this relationship context between AI entities and humans.

---

[193] Cofone, *supra* note 192, at 174.

[194] *Id.* at 174–75.

[195] *Id.* at 195–96.

[196] *See generally* Kate Darling, *Extending Legal Protections to Social Robots: The Effects of Anthropomorphism, Empathy, and Violent Behavior Towards Robotic Objects*, ROBOT LAW 213–31 (2014) (discussing the issue of robots that are designed to interact with humans on a social level).

[197] *See* Cofone, *supra* note 192, 175–76. Scholars attempt to establish an analogy between the owners of AI entities, such as autonomous vehicles, and owners of animals, such as dogs. Duffy & Hopkins, *supra* note 101, at 467. However, these attempts neglect many details. The mere fact that AI entities think and act independently from their human owners (to a variable degree) is a feature they share with animals. However, this common trait may not be enough to persuade us that dogs and AI machines are each other's peers. AI entities and dogs serve no common function, and the practical implications of their "usage" and "being" is very different. For example, AI entities are manufactured, while dogs are born like any other mammal. This difference is significant in the context of torts when determining who is responsible for AI-inflicted damages. In this scenario, the dog analogy falls short. A more suitable solution is necessary to reconcile our general principles with specific cases. *See supra* Part III, Section A.2.

[198] These frameworks mainly represent principal-agent relationships.

[199] Kenneth Coats, *Let the Robots Take Over: How the Future of AI Will Create More Jobs,* FORBES TECH. COUNCIL (Dec. 28, 2018), https://www.forbes.com/sites/forbestechcouncil/2018/12/28/let-the-robots-take-over-how-the-future-of-ai-will-create-more-jobs/#59808af23c6d [https://perma.cc/MM8X-5MXQ] (describing the use of robots in transportation, health care, home services, education, public safety, employment, and entertainment, among other areas of the economy).

As of now, in light of the social relationships cultivated with our AI companions,[200] the programmed goal assigned to them to complete on our behalf, and their proliferation among us, AI entities cannot be considered as adult humans yet, nor are they children or pets.

Agency offers us a way to utilize the many benefits of legal reasoning through analogies without, as cautioned by Judge Cardozo above, enslaving our thoughts.[201] On the contrary, it enables courts, officials, and lay people to draw parallels between new and old scenarios with a healthy dose of flexibility, which will prevent us from falling into the traps set up by legal analogies.[202] Moreover, the continuous back and forth reasoning process that stands at the heart of identifying and implementing the appropriate analogy also shields us from falling into these traps and enables us to verify that our general and specific principles will always reside in harmony.[203]

Treating AI entities as AI agents emphasizes the rejection of the homunculus idea evident in our previous discussion about electronic persons—that is, the notion that the AI entity has its own intentions, desires, and thoughts.[204] Those who created, owned, or dispatched the AI entity are responsible for its behavior, acts, and inflicted damages.[205] The AI entity is the agent of its creator, owner, or dispatcher and is not liable by itself.[206] In most AI-inflicted damages cases, this is the most adequate and best-suited analogy through which we can identify the liable party.

D. Analogies of the AI Industry

This sub-chapter explores a different set of analogies to the AI market as a whole. These analogies try to push for the opposite of a strict liability regime, which was embedded in the analogies presented thus far, in the form of a no-liability regime. The main underlaying justification for this liability regime rests on the stifling of AI innovation. Reviewing these analogies will present how they are inadequate to address our subject matter and how they direct us to the appropriate path of strict liability via agency, rather than a no-liability regime. These analogies try to establish a connection between the AI industry and (1)

---

[200] Calo, *supra* note 192, at 532 (stating robots today have "social valance," in that they feel more like living beings); Balkin, *supra* note 3, at 46 (stating people may substitute AI agents and robots for other living things, including humans, in certain contexts).

[201] Berkey v. Third Ave. Ry. Co., 244 N.Y. 84, 94 (1926).

[202] Cofone, *supra* note 192, at 174–77.

[203] *Id.*

[204] Balkin, *supra* note 17, at 1223 (referring to the "homunculus fallacy" – the idea there is a person inside the machine who has good or bad intentions and makes the machine do good or bad actions).

[205] Balkin, *supra* note 3, at 52.

[206] *Id.*

the aviation industry system of time limits and predictable pay-outs; and (2) the vaccine system and regulation, which includes a mass compensation fund in case damages happen.[207]

1. The Aviation Industry

It seems logical *prima facie* to compare an AI entity and an airplane flying on autopilot mode while a pilot still sits at the pilot's chair.[208] In *Brouse v. United States*, which was decided in 1949, the Ohio District Court placed liability on a pilot, who was using autopilot mode, for a crash that happened between two airplanes.[209] The court stated people in charge of airplanes under robotic control must keep a "proper and constant lookout," particularly where the airplane is in flight across recognized and established airways.[210] However, this case is inadequate when it is applied to the AI market in three ways.

First, there is no assurance that there will even be a human driver or operator in future autonomous car models and other AI entities. Therefore, placing liability on the operator will become irrelevant if she has no role in the actual operation of the AI-based machine.[211] Second, the person who was liable for the accident was an army airplane pilot flying a plane owned by the government.[212] This may have contributed to the conclusion the pilot was liable because the government has deep pockets and could pay for the damage.[213] The same policy logic will usually not apply in the commercial AI context where governmental bodies are rarely a crucial part of the manufacturing and marketing ecosystems.[214]

Third, there is an immense difference in operating an airplane in the sky and operating any other "robotic control" device on the ground or at sea. The airplane industry was built upon the usage of autopilots, and it is a well-accepted custom to use them in our aerial space.[215] This space is much more rigorously regulated than any other physical space, especially since this case was adjudicated back in 1949. For example, the General Aviation Revitalization Act of 1994 granted immunity to manufactures of small planes and their parts from liability for eighteen

---

[207] Sheela Jayabala et al., *Artificial Intelligence Governance: A Heads up from Driverless Cars*, 34 WORLD APPL. SCI. J. 376, 379 (2016).

[208] Marchant & Lindor, *supra* note 80, at 1325.

[209] Brouse v. United States, 83 F. Supp. 373 (N.D. Ohio 1949).

[210] *Id.* at 374–75.

[211] *Id.*

[212] *Id.* at 374.

[213] Marchant & Lindor, *supra* note 80, at 1329 (noting the party with the "deep[est] pockets" is likely to be found liable in a crash involving an autonomous vehicle).

[214] The exception to this statement may be autonomous weapons machines.

[215] *See, e.g.,* Allison Hope, *How Autopilot on Planes Works,* CONDE NAST TRAVELER (Aug. 7, 2017), https://www.cntraveler.com/story/how-autopilot-on-planes-works [https://perma.cc/U6VG-Y9EF] (describing the history and use of autopilot).

years in order to make sure the industry will survive notwithstanding liability suits.[216] The net social benefit of having an airline industry motivated Congress to act for its protection.[217] Whether we can expect Congress to do the same in the AI context is highly doubtful due to the vast implications AI entities have in various fields, unlike airplanes which are restricted to the field of aviation alone. It will be difficult to grant complete immunity to specific manufactures, as will become clear in our discussion about the vaccination industry below.

Time limits and predictable pay-outs are an important tool at the hands of the aerial industry; however, they are less apt in the AI context. Assuming AI entities are like autonomous vehicles, time limits are inapplicable to AI entities because the notion of applying an identical rigorous time limitations scheme to the usage of autonomous vehicles on the ground or autonomous vessels at sea is unliberal and impractical. The demands for certifications are vastly diverse and less rigorous when we depart the aerial space, and the physical infrastructures required for an efficient participation of a vast number of vehicles on the ground are extremely different than that required for airplanes in the air.[218] The pay-out approach may be applicable to autonomous entities on the ground in the form of strict liability with a built-in roof limitation.[219]

These elements lead us away from placing liability on the pilot of a plane flown by autopilot, as was decided in *Brouse*. Unless the operator of the AI entity has committed an act that completely deflects the damage endured from the AI entity itself to her behavior, it will be difficult to claim she should be held liable over the manufacturer of the AI device. We can see a contemporary example for this notion in the Uber autonomous vehicle accident which lead to the death of Elena Herzberg, a pedestrian in Tempe, Arizona, in March 2018 while a human driver was sitting in the driver's seat.[220] The investigation of the accident discovered the car recognized the pedestrian prior to the collision but did not take any active measures to stop the car or alert the

---

[216] General Aviation Revitalization Act, 49 U.S.C. § 40101 (2019).

[217] *See* Victor E. Schwartz & Leah Lorber, *The General Aviation Revitalization Act: How Rational Civil Justice Reform Revitalized an Industry,* 67 J. AIR L. & COM. 1269, 1341 (2002) (stating there was social policy that supported the General Aviation Revitalization Act).

[218] In other words, comparing the possible and desired levels of activities of passenger vehicles versus airplanes is like comparing apples and oranges.

[219] *See, e.g.*, Abraham & Rabin, *supra* note 80. This will probably require a combination of a strict liability regime with an insurance policy, likely a limited one, that will enable the usage of AI entities despite their safety issues. *Id.*

[220] Troy Griggs & Daisuke Wakabayashi, *How a Self-Driving Uber Killed a Pedestrian in Arizona*, THE N.Y. TIMES (Mar. 21, 2018), www.nytimes.com/interactive/2018/03/20/us/self-driving-uber-pedestrian-killed.html [https://perma.cc/J7M2-GV59].

driver.[221] The safety driver, Rafaela Vasquez, might be facing charges of vehicular manslaughter, though it is not clear yet whether the police will indeed press charges.[222] However, it is extremely difficult to expect Rafaela Vasquez, or any other person in a similar position, to be able to react in time to such an incident. This situation seems much like security guards at stores and shopping malls who spend their day in a mind-numbing routine and are expected to spring into action upon a split-second notice in case of danger.[223] Together, these considerations suggest that it is more difficult to place liability on the operator of the machine rather than the designer or manufacturer, when the former's main rule is merely to observe the AI entity's behavior rather than dictate or guide it.[224]

2. The Vaccination Industry

Those who advocate for comparing the AI industry to the vaccination industry claim that, much like vaccines, AI entities offer a net safety gain to society as a whole.[225] The social value of vaccines is generally undisputed, but they were exposed to many lawsuits before federal preemption laws were put in place to save the vaccine industry from bankruptcy.[226] After an $8.5 million verdict was granted in favor of an injured party who contracted polio from an oral polio vaccine,[227] Congress intervened and legislated the National Childhood Vaccine Injury Act of 1986 which established a system of regulations and standards for vaccines that limited the liability of vaccine suppliers and established a fund from which money could be taken in case of injury.[228]

As in the aviation context, granting immunity via a no-liability regime or a limited liability regime, created by legislative protections, will surely spur innovation. However, it will come at the price of "diminish[ing], if not eliminat[ing], the incentives for manufacturers to

---

[221] Chaim Gartenberg, *Safety Driver of Fatal Self-Driving Uber Crash was Reportedly Watching Hulu at Time of Accident*, THE VERGE (Jun. 22, 2018), www.theverge.com/2018/6/22/17492320/safety-driver-self-driving-uber-crash-hulu-police-report [https://perma.cc/L4JV-N55K].

[222] *Id.*

[223] Selbst, *supra* note 80, at 30; Gadi Perl, *Several Thoughts Following the Fatal Uber Accident in Tempe*, THE FEDERMANN CYBER SECURITY CENTER – CYBER LAW PROGRAM (Apr. 11, 2018), www.csrcl.huji.ac.il/people/several-thoughts-following-uber-fatal-accident-tempe [https://perma.cc/CS4T-YYP9].

[224] Dana Hull, Mark Bergen, & Gabrielle Coppola, *Uber Crash Highlights Odd Job: Autonomous Vehicle Safety Driver*, BLOOMBERG (Mar. 23, 2018), www.bloomberg.com/news/articles/2018-03-23/uber-crash-highlights-odd-job-autonomous-vehicle-safety-driver [https://perma.cc/W4L7-HUR9].

[225] Weston Kowert, *The Foreseeability of Human-Artificial Intelligence Interactions*, 96 TEX. L. REV. 181, 195 (2017).

[226] *Id.*; Marchant & Lindor, *supra* note 80, at 1331.

[227] Strong v. Am. Cyanamid Co., 261 S.W.3d 493, 521 (Mo. Ct. App. 2007).

[228] 42 U.S.C. §§ 300aa-1–300aa-34 (2019).

make marginal improvements in the safety of their products in order to prevent liability."[229] Marchant and Lindor believe it is preferable to endure this risk in order to spur AI innovation, specifically in the context of autonomous vehicles, sooner rather than later, even if it will cause damages, because even an incomplete autonomous vehicle will be safer than a human driver.[230] To them, the correct balance between the positive and negative incentives is a no-liability regime.[231]

On the other hand, there is no clear explanation for why the victim, by herself, should subsidize the damage she endured by these new technologies.[232] Furthermore, one can claim the mere advocation for a no-liability regime could actually hurt one's incentive to innovate.[233] This does not refer to the general idea of innovating something new, but rather to the constant refinement and enhancement of an existing technology that can be improved given the correct set of incentives to do so.

In *Bruesewitz v. Wyeth*, a parent of a minor who was allegedly injured from a diphtheria-tetanus-pertussis vaccine sued the manufacturer of the vaccine under a product liability claim.[234] The majority opinion, led by Justice Scalia, ruled against the parents and held the National Childhood Vaccine Injury Act preempts all claims against the manufacturer.[235] In her dissenting opinion, Justice Sotomayor noted the relationship between immunity and the lack of incentives to innovate.[236] She claimed the majority opinion "leaves a regulatory vacuum in which no one ensures that vaccine manufacturers adequately take account of scientific and technological advancements when designing or distributing their products."[237]

This may lead to the lack of internalization of the damages endured by a no-liability regime, which in turn will advance a failure to update the technology even when it is possible and desired.[238] This can be avoided by conditioning a party's non-liable status by mandating it implement improvements when possible.[239] However, this solution is probably economically impractical due to lack of internalization. Because there is no punishment or accountability, this means there is no incentive to improve.[240] Society must act to protect activities that are

---

[229] Marchant & Lindor, *supra* note 80, at 1337.
[230] *Id.* at 1340.
[231] *Id.*
[232] *See* CALABRESI, *supra* note 146, at 116–17.
[233] Kowert, *supra* note 225, at 199.
[234] Bruesewitz v. Wyeth L.L.C., 131 S. Ct. 1068 (2011).
[235] *Id.* at 1082.
[236] *Id.* at 1098 (Sotomayor, J., dissenting).
[237] *Id.* at 1086.
[238] *See* Kowert, *supra* note 225, at 199.
[239] *See id.*
[240] *Id.*

beneficial to us as a whole but that cannot be at the expense of injured victims who are left with no remedy.[241] The existence of a fund providing compensation to the injured party may realize the goals of corrective justice and fairness to some degree, but it does not achieve the goals of deterrence and efficient risk allocation.

The airplane industry and the vaccination industry analogies, which represent to a different degree the no-liability regime, diverge from the agency analogy, as evident from the differences in the liability regimes that result from each of them. The government views these two industries as having high value for society as a whole.[242] In the vaccination context, the government has detached the inherent connection between a manufacturer and its products for a cause it considers as worthy at all cost.[243] Whether we agree with this decision is a question of policy rather than fact. The rationale that stands behind the decision to disengage the agency cord is understandable if we agree that vaccinations are beneficial to all.[244]

One can also understand this approach in the context of the aviation industry but to a lesser extent. Providing a grace period for a market we consider valuable to everyone in order to allow it to grow despite foreseeable damages is again a question of policy, in this specific case an economical one, which we may disagree with. It is true this immunity period allowed this market to blossom, at the expanse of human lives, but it blossomed, nonetheless. However, eventually it was decided that the benefits of a no-liability regime no longer justified the cost, and that airline companies needed to internalize and take responsibility for their damages in order to provide better service to their customers.[245] Today's technological advancements are far more progressive than those of 1994, when the General Aviation Revitalization Act was passed.[246] Providing a no-liability grace period to AI companies will probably not receive the same support the aviation industry did, which was controversial at best.[247] It may be offered with regards to small AI companies, as it was applied to small airplanes, in order to ensure "small players" can also take part in the race of AI development, even if they cause damages they cannot pay for.[248]

---

[241] *Id.* at 203–04.

[242] *Id.* at 196; Schwartz & Lorber, *supra* note 217, at 1341.

[243] Kowert, *supra* note 225, at 196–97.

[244] *See id.* at 196.

[245] Kerry V. Kovarik, *A Good Idea Stretched Too Far: Amending the General Aviation Revitalization Act to Mitigate Unintended Inequities*, 31 SEATTLE U. L. REV. 973, 985 (2008).

[246] *Id.* at 983.

[247] *See id.* at 974; *see also* Lawrence J. Truitt & Scott E. Tarry, *The Rise and Fall of General Aviation: Product Liability, Market Structure, and Technological Innovation*, 34 TRANS. J. 52 (1995).

[248] Kowert, *supra* note 225, at 198.

However, this should be dealt with via an anti-trust law channel, rather than that of torts, so injured parties will receive the remedy they need and deserve.[249] Our tolerance for innovation at the expense of victims who remain with no redress has reduced during the years, which lead to the adaptation of strict liability regime in the first place.[250]

A practical argument could be that, if anything, the grace period for AI companies should have already ended. For example, the autonomous vehicle has received lenient and convenient regulation from different states in the United States, allowing its creators to run experiments at ease with little to no responsibility since 2011.[251] These laws have already granted these manufactures a grace period which should have allowed these companies to improve their AI entities in a contained environment before they were certain these AI entities are safe enough to be tested in an uncontrolled environment. Therefore, if any sort of damage did occur, the principal in the chain of the AI entity development or operation should assume responsibility for their choice to use the AI entity in an uncontained environment.

E. Final Note—AI as Something Old or New?

AI entities may be treated as something completely new. This implies creating a new legal entity, doctrine, or category to manage or facilitate their existence in the fabric of society.[252] Such an approach would require time in order to fully understand and internalize the place of these AI entities in our world, which is constantly changing.[253] For this reason, adopting a completely new analogy to tackle the liability of AI entities seems less beneficial and more constraining than using an existing one. Since AI entities are evolving entities which may be treated as human beings in the future, focusing on discovering a new and improved analogy may derail our efforts to create a comprehensible and coherent tort regime for damages that occur now and will occur in the near future.

On the other side of this spectrum, one can see the damages inflicted by AI entities just like any other damage inflicted in an accident.

---

[249] Gary Myers, *The Different Treatment of Efficiency and Competition in Antitrust and Tortious Interference Law*, 77 MINN. L. REV. 1097, 1097–99 (1993) (explaining the differences and conflicts between recovery under antitrust laws and tortious interference laws).

[250] Jed Handelsman Shugerman, *The Floodgates of Strict Liability: Bursting Reservoirs and the Adoption of* Fletcher v. Rylands *in the Gilded Age*, 110 YALE L.J. 333, 377 (2000).

[251] Jayabala, *supra* note 207, at 378. Nevada began permitting the operation and testing of autonomous vehicles in 2011, followed by California and Florida. These states had flexible licensing requirements and left room for future standards to be developed. *Id.*

[252] Crootof, *supra* note 64, at 52 (suggesting that a new supplemental law is necessary to effectively regulate autonomous weapons systems).

[253] *Id.* at 58–59.

In this sense, an AI accident is not different from a car accident, a work accident, or a discriminatory employer. Thus, we should not rely on an analogy to anything different or new. Adding this layer of fiction, one can claim, does not contribute to the legal analysis. However, legal reasoning through analogy is not a fleeting method practiced by courts, officials, and scholars. It will be extremely difficult to uproot the usage of this method and not necessarily beneficial to the intersection of law and technology as a whole. Eventually, every new technology becomes old, but in order for it to be considered as such, it must first go through a process of acceptance by society and its institutions. Applying legal analogies is one of the ways we process its existence.[254]

There is a strong argument to be made that the use of legal analogies in the intersection of AI and tort law is not necessarily desirable or fruitful.[255] For example, analogizing AI entities to servants or products leads to the conclusion that the person behind AI entities are liable. Advocating for a straightforward strict liability regime achieves the same result, some may say, in a clearer, less fanciful way.

Legal analogies have limitations and may inhibit our ability to properly regulate AI entities if we continue to view them as something they are not.[256] Nonetheless, legal analogies can help reveal the most equitable underlying principle to govern AI liability. They transform the unintuitive into a common language we can understand.[257]

Judges do not have to be experts in specific fields; however, they must possess tools and instruments to enable them to analyze the scenario in front of them despite this lack of expertise.[258] Legal analogies supply these tools and create a legal language all can share and understand.[259] They help bridge the gap created by this deficiency and provide the court, the legislature, and executive officials with the appropriate tools and instruments they need to utilize legal reasoning by analogy.[260]

Throughout this section, the back and forth voyage has been put into the test in our search for the desired equilibrium point. This section reviewed the possible analogies, the values that stand behind them and their application in specific cases. This section demonstrates that many suggestions do not reach the equilibrium point this Article wishes to attain. Their specific application does not fulfill the general principle

---

[254] *See supra* Part II, Section C.

[255] *See generally* Crootof, *supra* note 64 (stating that when there is no appropriate legal analogy, a new analogy or set of rules must be created).

[256] *See supra* Part II, Section B.

[257] Blavin & Cohen, *supra* note 48, at 267 (stating "[w]hen courts encounter new technologies not yet anticipated by the law, their reliance on analogical reasoning plays a profoundly important role in the application of proper legal rules.").

[258] *See* AHARON BARAK, THE JUDGE IN A DEMOCRACY 101–07 (2006).

[259] Blavin & Cohen, *supra* note 48, at 266–67.

[260] *See id.* at 266 (stating that metaphors can be used to answer questions about new social and legal problems).

which stands behind the usage of the analogy and they cannot be adjusted by their mutual influence without undermining themselves. As of now, it seems that only the analogy of agency reaches harmony between the general principles and specific instances of AI entities inflicting harm. Our voyage has led us to our warranted discussion about the legal analogy of AI agent-servant and its application to AI entities' liability.

IV. TREATING AI ENTITIES AS AGENTS-SERVANTS

On the one hand, the AI agent-servant analogy is the most appropriate approach to the analysis of AI entities. On the other hand, there is something inherently false about this analogy as AI entities are not real servants, and they were never created to be as such.[261] The true application of this servant analogy is less about the similarities of an agent and an AI entity in a deep sense, and more about the legal meaning that stands behind it—that is, the direct applicability of strict liability with regards to the harmful actions of a servant under the control and care of its principal.[262]

The reasons and values, which stand behind this normative choice, states that there are scenarios in which it is appropriate to treat AI users as more than just property owners. They may hold the ability and responsibility to do more in order to prevent damages and harms than mere product owners. Given this, our regulatory structure should create incentives to make sure these AI users and manufacturers do so. A strict liability regime aligned with the *respondeat superior* doctrine will enable this to happen. Furthermore, a servant, which is a sub-category of an agent, is unique in that it acts under the direct control and supervision of its master, unlike other forms of agency which have greater discretion to make independent decisions.[263]

---

[261] An exception to this is servant-robots. *See, e.g.*, Evan Ackerman, *Care-O-Bot 4 is the Robot Servant We All Want but Probably Can't Afford*, IEEE SPECTRUM (Jan. 29, 2015), www.spectrum.ieee.org/automaton/robotics/home-robots/care-o-bot-4-mobile-manipulator [https://perma.cc/DSB2-DK78] (describing a robot that can help with household tasks); Evan Selinger, *Robot Servants are Going to Make Your Life Easy, Then They'll Ruin it*, WIRED (Sept. 5, 2014), www.wired.com/2014/09/robot-servants-are-going-to-make-your-life-easy-then-theyll-ruin-it/ [https://perma.cc/LD5S-QG38] (describing a "Jetsons"-style "family robot").

[262] *See* Kerl v. Rasmussen, 682 N.W.2d 328, 331 (Wis. 2004) ("Vicarious liability under the doctrine of respondeat superior depends upon the existence of a master/servant agency relationship.").

[263] CRAIG B. GLIDDEN, 4 WEST'S TEXAS FORMS, BUSINESS LITIGATION Ch. 15 Introduction (2d ed. 2019 update). In some instances, courts have held that there is no legal difference between an agent and a servant. *See, e.g.,* Murray v. Hills Cab Co., 198 N.E.2d 466 (Ohio Ct. App. 1963), but it seems to have some factual importance in the way we describe the nature of a relationship.

The *Restatement (Second) of Agency* sets a few criteria to help determine "who is a servant."[264] These include, *inter alia*,

(a) the extent of control which . . . the master may exercise over the details of the work; . . . (c) the kind of occupation with reference to whether, in the locality, the work is usually done under the direction of the employer . . . (d) the skill required in the particular occupation; . . . (f) the length of time . . . (i) whether or not the parties believe they are creating the relationship of master and servant.[265]

The AI agent fulfils the vast majority of these criteria in its relationship with its human principal, given its unique features and lack of independence in choosing tasks. For example, the human principal has great control over the assignments of her AI agent. Given the nature of the agent, its actions are carried out under the direction and discretion of the principal, and the AI agent is specifically crafted with a particular set of skills to accomplish specific assignments given to it by its principal [266] for as long as the AI agent is operational.

Of course, the analogy is not perfect, and it raises a difficult set of questions, the most important one is who should be viewed as the principal in the relationship with an AI entity. Is it the manufacturer of the AI entity, the humans who purchased the AI entity, a third party who loaned the AI entity from its owner or creator (much like a borrowed car or tool), or maybe someone else entirely? What follows is a discussion of these questions and others about the application of the agency theory on AI entities and its direct connection to enforcing a strict liability regime when damages occur.

A. The Meaning and Purpose of the Term "Agent"

Even scholars that advocate for legal analogies other than agency, or for no specific analogy at all, often use the term "AI agents."[267] This indicates the existence of a gap, or at the very least inconsistency, in the usage of the term in the legal context.

The *Restatement (Third) of Agency* defines agency as:

> the fiduciary relationship that arises when one person (a 'principal') manifests assent to another person (an 'agent') that the agent shall act on the principal's behalf and subject to

[264] RESTATEMENT (SECOND) OF AGENCY § 220(2) (AM. LAW INST. 1958).
[265] *Id.*
[266] This is a main feature of narrow AI, which is the prevalent type in today's market. *See* Tannya D. Jajal, *Distinguishing Between Narrow AI, General AI and Super AI*, MEDIUM (May 20, 2018), www.medium.com/@tjajal/distinguishing-between-narrow-ai-general-ai-and-super-ai-a4bc44172e22 [https://perma.cc/CR5U-NXDK] (describing narrow AI as that which is capable of performing a single specific task at any given time).
[267] *See, e.g.,* CHOPRA & WHITE, *supra* note 103; Balkin, *supra* note 3; Cofone, *supra* note 192.

> the principal's control, and the agent manifests assent or otherwise consents so to act.[268]

Black's law dictionary defines agency as "[a] fiduciary relationship created by express or implied contract or by law, in which one party (the *agent*) may act on behalf of another party (the *principal*) and bind that other party by words or actions."[269] But outside of the realm of law, and sometimes even within it, agency can take on a different form and refer to an organization, company, or bureau that provides service.[270] Munday states the "word 'agent' is often 'used indiscriminately to describe individuals and entities whose activities, in strict legal terms, are not actually governed by the law of agency.'"[271]

When the phrase *AI agents* is used, most scholars refer merely to an entity based on AI technology, which is capable of performing any basic action, such as being active. Similarly, the term "chemical agents," for example, refers to active chemical substances which possess the ability to hurt by the nature of their activity.[272] Others use the word agents, and agency in general, in the context of sentient beings who can think and have an agenda and purpose of their own.[273]

In contrast, this Article uses the term to describe AI entities as agents in a relationship with their fellow human principals—that is why thus far this Article described them in a generic non-anthropomorphism term by referring to them as "entities" in an attempt to differentiate the general features attributed to an AI entity. All of these uses are applicable and acceptable, but their usage must be more careful and mindful. They express very different concepts of the way AI entities should legally be treated and using the term interchangeably may lead to confusion in robotic and AI law. Furthermore, even if the term "agency" is used accurately and coherently within its legal sense in order to describe a legal relationship between an agent and a principal, it does

---

[268] RESTATEMENT (THIRD) OF AGENCY § 1.01 (AM. LAW INST. 2006).

[269] *Agency*, BLACK'S LAW DICTIONARY (7th ed. 1999).

[270] *Agency*, CAMBRIDGE DICTIONARY, http://www.dictionary.cambridge.org/us/dictionary/english/agency [https://perma.cc/344E-WJKA].

[271] RODERICK MUNDAY, AGENCY: LAW AND PRINCIPLES 1–2 (3rd ed., 2016). Munday goes on to quote Lord Herschell, stating that "[n]o word is more commonly and constantly abused than the word 'agent.'" *Id.* at 2 (quoting Kennedy v. De Trafford [1897] AC 180, 188).

[272] *What is a Chemical Weapon?*, ORGANIZATION FOR THE PROHIBITION OF CHEMICAL WEAPONS, https://www.opcw.org/sites/default/files/documents/Fact_Sheets/English/Fact_Sheet_4_-_CW_types.pdf [https://perma.cc/GR5G-7CGS] (stating that "[t]he toxic component of a chemical weapon is called 'chemical agent.'").

[273] *See* Cofone, *supra* note 192, at 183 (stating that since A.I. robots lack agency, they cannot effectively be held civilly or criminally liable for their actions).

not necessarily mean the same narrow rules will apply to different types of AI agents.[274]

B. Who is the Principal?

The question of identifying and naming the principal is not a simple one to answer, due to the unique structure of the AI industry and the many human hands that are a part of the development, manufacturing, training, and operation of an AI entity. It may be the case that a specific AI agent will have multiple principals over the course of its existence, and it is important to identify the appropriate one, or ones, in a given situation.

1. The Possible Approaches to Identify the Principal

To name the appropriate principal in case of an accident, two possible approaches should be considered. First, analyze the specific task at hand that the AI entity is in charge of completing and who is the person who sent it to complete that specific task.[275] Second, observe who has the highest capacity and capability to affect the actions of an AI entity in means of monitor, supervision, and guidance.[276]

A good example for the first approach can be found in the Israeli agency legislation, that is its Agency (*Shlicot*) Law.[277] This law states that the agent-principal relationship applies on every action which is reasonably necessary for the proper execution of the subject of the mission, when there is no limitation on the authorization given to the agent.[278] Furthermore, the law states that an agent is permitted to carry out any urgent and unpredictable action which is reasonably required in order to safeguard the matters of the sender (i.e., principal), which is related to the subject of the mission, even if these actions exceed its authorization limitations.[279] The emphasis here is on the mission given to the agent by its principal and the acts the former had to carry out in order to fulfil that specific mission. If, in the course of conducting the

---

[274] MUNDAY, *supra* note 271, at 3 ("[S]pecific rules and customs may apply to different species of agent.").

[275] *See, e.g.,* Agency Law, 5725-1965, 19 L.S.I. 231§ 1(a) (Isr.) [hereinafter Israeli Agency Law] (stating that agency is the grant of power to act in the place or name of the principal).

[276] *See* RESTATEMENT (THIRD) OF AGENCY § 1.01, cmt. e (AM. LAW INST. 2006) (stating that the principal instructs and controls the agent).

[277] Israeli Agency Law, *supra* note 275.

[278] *Id.* at § 5(a) (stating that "[a]gency extends—unless limited by the authorisation—to any act reasonably required for the proper carrying out of its object").

[279] *Id.* at § 5(b) (stating that "[a]n agent may do any urgent and unforeseen act reasonably required for safeguarding the interests of the principal in connection with the object of the agency even if such act exceeds the scope of the authorisation.").

mission, a harm was inflicted, the principal who sent the agent on the mission is liable.[280]

The second approach is the American one, according to the definition of agency in the *Restatement (Third) of Agency*.[281] This approach takes a broader look at the situation and examines the correlation between the principal and agent, as well as the former's ability to control and guide the latter.[282] The focus is on the nature of the relationship between the agent and principal rather than on the specific mission at hand.[283] It seems this approach is more appropriate in the AI context because the AI entity is not necessarily an agent sent to complete a specific mission or task, but rather it is in a constant state of "being an agent" for the benefit of others, mainly of its principal, similar to the behavior of a servant.

Moreover, this approach focuses on the question of where is the greatest pressure point for who is willing and able to take into consideration the costs its AI agent may inflict. Based on these costs, that person—that is, the principal—is in the best position to make a decision whether to better equip and train the AI agent, or to pay the price in the form of a monetary sanction or an insurance premium. Given the relatively low number of options for identifying the principal, the administrative costs associated with naming her should not exceed the administrative costs associated with adjudicating each case as a negligence claim. Thus, this approach seems to be the most efficient in the AI liability context.[284]

When we discuss human servants, we can roughly divide their lives into two separate parts: the agent share and the autonomous share. While the former refers to performing the tasks assigned by the principal, the latter refers to anything in that servant's life which is not related to her agency role and is not under the control or guidance of her principal. This separation does not exist when we discuss AI entities. They do not possess an autonomous part of their existence, but only a mono-purpose aspect. A human servant's main mission is to complete the orders her principal provides; however, they eventually have a sperate part of existence where they are in charge of and are responsible for their own actions, decisions and mistakes, outside the realm of their

---

[280] *See infra* Part IV, Section B.3 (explaining that in the AI context, the idea that a principal is responsible for an agent's actions is not applicable with relation to intentional torts conducted by the agent).

[281] *See* RESTATEMENT (THIRD) OF AGENCY, § 1.01 (AM. LAW INST. 2006). The Israeli and American approaches were examined because they best represent the continuum between the two possible approaches within the common law system.

[282] *See* Susan P. Shapiro, *Agency Theory*, 31 ANN. REV. SOC. 263, 273 (2005) (stating that principal control is critical in the law of agency).

[283] See RESTATEMENT (THIRD) OF AGENCY § 1.01 cmt. d (stating that agency is a consensual relationship).

[284] *See* CALABRESI, *supra* note 146, at 225 (expanding on administrative costs).

missions as servants. AI entities do not "enjoy" this luxury. They are agents from the moment they were switched on (creation of agency) until the moment they will be switched off (termination of agency).[285] Given this, no specific mission dictates their behavior, but rather the ongoing services they provide to their principals as their ongoing servants.

2. Creation and Termination of an AI agency

Agency is created by contract (i.e., consent of the parties to enter such a relationship even if no written contract was signed) or is created as a matter of law (i.e., implied from circumstances).[286] There are no rigid and formal requirements for the creation of an agency, which supports the notion this relationship can be established with AI entities on the agent end. [287] The mere existence of AI entities and their affiliation with their principal leads to the application of agency, even if the agent itself has not given its explicit consent to this relationship, an already problematic concept in the AI realm.[288] Unlike traditional agency, the AI entity's sole purpose is its role as an agent. Therefore, the creation of an agency starts from the creation of the AI entity itself or, rather, its activation. This directly connects to the mono-purpose agenda of the AI entity presented earlier, which differentiates traditional agency from AI agency.

The AI agent can only be terminated if the principal chooses to end the agency, because even if it could choose, its choice is really only one it was already programmed for.[289] Either the AI entity is out of commission or is reassigned to another principal (either by choice or death of the former principal), thus only terminating a specific agency relationship of the AI entity out of a continuum of agencies.

---

[285] While AI may not yet have the luxury of being in charge of itself, there are arguments for creating AI in such a way that it has free will. *See* Alzbeta Krausova & Hananel Hazan, *Creating Free Will in Artificial Intelligence*, *in* Proceedings of the International Conference Beyond AI 96, 107 (2013) (arguing that artificial intelligence should include free will but that it would be highly complex and the results would be uncertain); *see also* Cindy Van Rossum, Liability of Robots: Legal Responsibility in Cases of Errors or Malfunctioning (2018) (unpublished L.L.M. paper, Ghent University), https://lib.ugent.be/fulltxt/RUG01/002/479/449/RUG01-002479449_2018_0001_AC.pdf [https://perma.cc/4VWA-G6XD] (noting that one hundred percent autonomous robots do not exist yet).

[286] MUNDAY, *supra* note 271, at 15–16.

[287] *Id.* at 12; HOWARD BENNETT, PRINCIPLES OF THE LAW OF AGENCY 2 (2013).

[288] CHOPRA & WHITE, *supra* note 103, at 29. The notion of robot consent is also raised in the context of whether sex robots can consent. *See generally* Frank & Nyholm, *supra* note 145; Laura Bates, *The Trouble with Sex Robots*, THE N.Y. TIMES (July 17, 2017), www.nytimes.com/2017/07/17/opinion/sex-robots-consent.html [https://perma.cc/6T3F-BVLR].

[289] Rossum, *supra* note 285, at 18 (stating that even robots which decide to do something are actually responding to situations for which they were programmed).

However, the human agent has other options to terminate the relationship which are not available to an AI agent. For example, the agent can choose to terminate it,[290] the agreed period of the agency can come to an end, or the agent can accomplish her allotted task.[291] The latter two are inapplicable in the AI agent context, due to the specific-purpose feature of AI agents, which is unlimited by time or specificity of a task. Termination by operation of law can also come into effect if the principal or agent dies, the principal becomes mentally incapacitated, or the principal becomes insolvent.[292] All of these may very well happen in the case of an AI agent but that only terminates a specific agency within the life span of the AI entity. The death, insolvency, or mental incapacitation of the principal will only lead to a change in the identities of the potential principals, not to the end of the AI entity's role as an agent.

These unique features of an AI entity, mostly its mono-purpose agenda which substantially effects the creation and termination of AI agency, strengthen our conclusion that the second approach, which focuses on the relationship of the agent (servant) principal rather than on a specific mission, is a more appropriate approach to take to find the specific principal that is in control of the AI servant in a given situation, and thus liable for its indiscretions. This approach suits the mono-purpose feature—the focus is on the AI agent and its correlation to possible interchanging human principals. The ending of one relationship does not cancel the grander scheme of the AI's purpose to serve, thus it will help us identify the appropriate principal by focusing on the AI agent itself and on its obligations to its principal, rather than on its mission.

3. Fiduciary Duties

Another aspect of the agent-principal relationship is the obligations the agent owes the principal (e.g., fiduciary duties), vis-à-vis the control and guidance of the principal over its agent.[293] Identifying who owes fiduciary duties may also be used once we are trying to identify the appropriate principal. However, this index is harder to measure and utilize in order to identify the principal because the AI entities are not programmed to act in a specific way towards its principal. An AI entity usually treats its human surroundings in the same manner, unless it is programmed otherwise. Thus, pointing to a specific human as its principal based on the duties it owes her may prove to be a confusing measurement in our search for the appropriate principal and an

[290] This choice may no longer be reserved to humans if AI entities are eventually able to express and execute free choice.
[291] MUNDAY, *supra* note 271, at 34.
[292] *Id.* at 14.
[293] BENNETT, *supra* note 287, at 85; MUNDAY, *supra* note 271, at 11.

inherent limitation of the analogy of agency. If it was specifically programmed to owe fiduciary duties to a specific entity (i.e., programmed to be loyal), we need to question who decided to program it this way and whether it is really an indication of the appropriate principal. Otherwise, it is of no added value in helping us to identify the principal.

The concept of fiduciary duties is meant to prevent, control, and minimize the intrinsic conflict of interests that lays in the basis of the relationship between a human agent and a human principal.[294] This conflict rises once the interest of a human agent diverges from that of the principal in a manner that leads the agent to act on behalf of her personal interests, rather than those of her principal.[295] These are not common in the AI context. AI agents do not have their own personal interests but rather only those which were assigned to them as part of their mono-purpose agenda. The problem that may arise once an AI agent is dispatched is not that of conflicting interest and loyalty but rather excessive loyalty.[296] This may lead the AI agent to take bizarre or extreme measures in order to achieve the assigned task and, in the process, can cause damages for the sake of the assignment. This principle is also relevant in the context of intentional torts which are not applicable to AI agents because their actions will always be for the purpose of promoting the agenda of another person—their principal.[297]

The fact that we cannot guarantee the existence of fiduciary relationships between the AI agent and the human principal does not undermine the agency analogy because there is no conflict of interest between the agent and the principal to justify the existence of a fiduciary

---

[294] *See* Robert Flannigan, *The Economics of Fiduciary Accountability*, 32 DEL. J. CORP. L. 393, 420 (2007) (stating that fiduciary duties can prevent "mischief").

[295] *Definition of Principal/Agent Problem*, FINANCIAL TIMES, http://markets.ft.com/research/Lexicon/Term?term=principal%2Fagent-problem [https://perma.cc/U7EZ-SF37] ("Agency problems arise when the incentives between the agent and the principal are not perfectly aligned and conflicts of interest arise. As a result, the agent may be tempted to act in his or her own interest rather than the principal's. Conflicts of interest are almost inevitable.").

[296] *See, e.g.,* Curtis E.A. Karnow, *The Application of Traditional Tort Theory to Embodied Machine Intelligence*, *in* ROBOT L. 1, 1 (Ryan Calo, et al. eds. 2013) (stating "[a]sk a humanoid robot to reach his right hand to touch his left ear. In most of the cases I saw – the robot tried to get to the left ear through the head.").

[297] Once a human agent commits an intentional tort, she is solely liable for her behavior and the principal will not be held liable. There are three exceptions to this rule: if the conduct was (1) specifically authorized by the principal; (2) natural from the nature of the employment; (3) motivated by a desire to serve the principal. All of these exceptions are applicable in the AI context given the agency relationship. *See* Burlington Indus., Inc. v. Ellerth, 524 U.S. 742, 756 (1998); Doe v. Forrest, 853 A.2d 48, 67 (2004); Dianne Rosky, *Respondeat Inferior: Determining the United States' Liability for the Intentional Torts of Federal Law Enforcement Officials*, 36 U.C. DAVIS L. REV. 895, 903–04 (2003); s*ee also* PAGALLO, *supra* note 3, at 120 (elaborating on intentional torts in the AI context).

relationship in the first place. This lack of fiduciary duties also shows the existing gap between the agency analogy in the human context and that in the AI context. This gap emphasizes the importance of the strict liability regime that stands behind the agency analogy and its applicability in the AI context, rather than the perfect adequacy of the analogy itself.[298]

4. The Problem with Identifying the Principal

While discussing Rawls's reflective equilibrium and its utilization in finding the appropriate analogy for AI entities, a productive-manageable problem arose in the form of identifying the appropriate principal or principals in a given situation.[299] An *ex-ante* approach to regulating the identity of the principal in advance is not a productive path to take, despite its benefit in minimizing administrative costs, due to various degrees of control and guidance each entity has under different circumstances which cannot be dictated in advance.[300] Deciding the identity of the principal in advance, with no consideration to a given situation, disrupts the meaning of the agency analogy and the back and forth reasoning that leads us to it and will continue to develop it in the future. In order to find the best-suited principal, the pendulum must swing back and forth until it finds its equilibrium position.[301] Forcing it to stay in one place will prevent us from reaching that much desired equilibrium.[302] No single across-the-board rule can be decided on and applied with regards to a predefined entity. Each case should be examined separately to identify the principal given the agency relationship.

The identity of the principal will change per instance and will heavily depend on the circumstances of the accident. This will include consideration of factors, such as the level of involvement, supervision, monitoring, and ability to direct the actions of the AI agent in light of the damage that ensued.[303] At the early stages of an AI agent development, this level of control will be attributed more frequently to

---

[298] *See* Vladeck, *supra* note 91, at 146 (elaborating on strict liability in the AI context); *see also* Steven Shavell, *On the Redesign of Accident Liability for the World of Autonomous Vehicles*, (NBER Working Paper No. 26220, 2019), http://www.nber.org/papers/w26220 [https://perma.cc/95G6-DRN5] (suggesting a new form of strict liability in an autonomous vehicles context).

[299] *See supra* Part II, Section A.

[300] See CALABRESI, *supra* note 146; HALLEVY, *supra* note 74, at 185–229 (explaining the sanctioning of AI entities in the criminal context).

[301] *See* RAWLS, *supra* note 7, at 48 (explaining how going back and forth allows us to find an accurate description of the initial situation).

[302] *Id.* at 76 (arguing that it is a mistake to focus on a single transaction viewed in isolation).

[303] This Article has discussed several possible principals: owner, operator, designer, trainer, programmer, and user. As the AI entity progresses, each potential principal will become more or less involved due to their role and job duties.

the designer, programmer, trainer, or manufacture of the AI agent rather than its operator or owner.[304] The more the usage of these AI agents becomes pervasive, the more likely the operator's or owner's level of control and monitoring will result in identifying her as the appropriate principal.

Respondeat superior does not require the principal to be at fault, but rather that an agent performed a task that was within the realm of authority given to it by their principal.[305] Our ability to properly identify the correct principal according to this analogy will evolve over time and depend on the connections of control and supervision the AI entity has with its environment.

An interestingly complex scenario can arise if an AI agent inflicts damages because it was hacked and controlled from afar by someone other than those identified as the potential principals.[306] In this case, the hacker is the ad-hoc principal for the damages, but in most cases, it will be difficult to locate and prosecute her. However, the hacking itself can be seen as the tortious act that lead to the damages. Therefore, it will be difficult to hold the operator or owner liable given their low ability of control to prevent hacking. The designer, programmer, or manufacturer should be seen as the more appropriate principal in this scenario and therefore the liable party.[307]

5. Multiple Principals

There will be cases where an AI agent inflicts damages while it is under the control and guidance of more than one principal. These two

---

[304] *See supra* Part III, Section D (discussing the Uber autonomous vehicle accident).

[305] *See* Gary S. Green, *Respondeat Superior*, ENCYCLOPAEDIA BRITANNICA https://www.britannica.com/topic/respondeat-superior [https://perma.cc/GK8C-SH74] (stating that liability comes when an agent commits a crime while acting within the scope of the agent's authority and with an intent to benefit the corporation).

[306] *See* Alan Butler, *Products Liability and the Internet of (Insecure) Things: Should Manufacturers Be Liable for Damage Caused by Hacked Devices*, 50 U. MICH. J.L. REFORM 913, 914 (2017) (discussing the increasing incidence of connected devices being hacked); Mark A. Geistfeld, *A Roadmap for Autonomous Vehicles: State Tort Liability, Automobile Insurance, and Federal Safety Regulation*, 105 CALIF. L. REV. 1611, 1660 (2017) (discussing the potential for autonomous vehicles to be hacked and controlled by an unauthorized user); Evelyne Studer & Jacques De Werra, *Regulating Cybersecurity: What Civil Liability in Case of Cyber-Attacks?*, DROIT 511, 512–514 (2017) (discussing the question of who is liable when a device is hacked and examining the repercussions of these types of hacks); John G. Browning & Shawn Tuma, *If Your Heart Skips a Beat, It May Have Been Hacked: Cybersecurity Concerns with Implanted Medical Devices*, 67 S. C. L. REV. 637, 638 (2016) (discussing the potential for medical devices to be hacked).

[307] This short discussion is not exhaustive with regard to the process of identifying the appropriate principal in a hacking situation. It is only meant as an anecdote to present the complexity of the situation and the unpredictable parties that may be added to the potential principals.

or more principals are referred to as "multiple principals" and can be viewed as joint principals who are in co-control over the AI agent's activities.[308]

Article 41 of the *Restatement (Second) of Agency* discusses this issue of several principals under the title "interpretation of authority where principals or agents are joint."[309] It states in sub-section (1) that "[u]nless otherwise agreed, authority given by two or more principals jointly includes only authority to act for their joint account."[310] This notion repeats in article 3.16 of the *Restatement (Third) of Agency*, which states "[t]wo or more persons may as coprincipals appoint an agent to act for them in the same transaction or matter."[311] Article 17(b) of the Israeli agency legislation also deals with this situation. It states that "[w]here one authorization is given by several principals, it shall be presumed that the agent is to act by the joint authority of all of them."[312] The Louisiana Civil Code states in Section 3015 that "[m]ultiple principals for an affair common to them are solidarity bound to their mandatary."[313] This legislation presents multiple principals as an issue that is handled via joint and several liability. This is also the case according to the French Civil Code.[314]

All principals have a role in the actions of the AI agent. This Article named several potential principals: owner, operator, designer, trainer, and programmer. When more than one entity can be identified as the AI's principal, all the relevant principals should be held liable for the damage that occurred jointly and severally.

This situation of multiple principals is a part of the above-mentioned productive problem that stands at the heart of the AI agency reflective equilibrium process: identifying the appropriate principal in a given case.[315] Finding the principal or principals is not an easy task. One AI agent can be under similar or different degrees of control from various principals.[316] Only by reasoning back and forth can one identify the appropriate entity to be held liable as the AI agent's principal.[317] An alternative is to legislate *ex-ante* the identity of the principal. In this

---

[308] Scherer, *supra* note 80, at 287.

[309] RESTATEMENT (SECOND) OF AGENCY § 41 (AM. LAW INST. 1958).

[310] *Id.*

[311] RESTATEMENT (THIRD) OF AGENCY § 3.16 (AM. LAW INST. 2006).

[312] *See* Israeli Agency Law, *supra* note 275.

[313] LA. CIV. CODE ANN. art. 3015 (2019).

[314] *See* CODE CIVIL art. 2002 (Fr.) ("Where an agent has been appointed by several persons, for a common affair, each of them is jointly and severally liable towards him for all the effects of the agency.").

[315] *See supra* Part IV, Section B.4.

[316] *See supra* note 303 (noting that some possible principals are: owner, operator, designer, trainer, programmer, and user).

[317] *See generally* RAWLS, *supra* note 7, at 48 (explaining how back and forth reasoning works).

context, as discussed above, given the identification process of the principal, this type of legislation seems inaccurate.[318] Naming one or more entities as constantly liable will fail to take advantage of the many benefits agency law has to offer via its principal's incentives, once we apply it on the relationship between AI agents and those who control them.

The structure of multiple principals is possible and recognized, as we can see from the various agency legislations, and it is an important tool to utilize in the AI entity context given the many human hands which have varying degrees of control and guidance over these agents.[319] The back and forth reasoning process will allow us to find the appropriate principal, as it led us to the appropriate legal analogy, it will eventually do the same with regards to the apt human principal itself.[320] This is a productive dilemma which is manageable by our legal system and should find its answer via our legal reasoning process, which ensures the principal will not be arbitrarily or rigorously chosen.[321]

Two examples will assist us in better understanding the application of agency and multiple principals when AI agents cause damages. The first involves an autonomous shopping mall guard robot running over an infant while patrolling the parking lot.[322] In the second case, a hiring algorithm selected only a specific segment of the population as worthy of hiring, and its results were adopted by a tech company.[323] These cases represent two types of tortious damages—physical harm to a person in the first example and discrimination and disparate impact in the second.

In order to simplify the first example, assume the owner and operator of the robot is the mall itself and not an outside company, and that they choose the path of the robot's patrol. Examining who has greater control and supervisory ability to monitor the actions of the robot will most likely lead us to name the mall as the principal, assuming it sent the robot to complete the task of patrolling and has the ability to monitor its actions from afar. However, it is possible the mall will not be named the sole principal of the robot if the designer or programmer of the robot was already aware of this malfunction and failed to update

---

[318] *See supra* Part IV, Section B.4.

[319] *See, e.g.*, RESTATEMENT (SECOND) OF AGENCY § 41 (AM. LAW INST. 1958) (defining authority in the context of joint principals or agents).

[320] *See* RAWLS, *supra* note 7, at 48.

[321] *Id.* at 5 ("[I]nstitutions are just when no arbitrary distinctions are made between persons in the assigning of basic rights and duties.")

[322] Steve Hoffer, *300-Pound Security Robot Runs Over Toddler at California Shopping Center*, HUFFPOST (July 13, 2016), www.huffingtonpost.com/entry/security-robot-toddler_us_57863670e4b03fc3ee4e8f3a [https://perma.cc/G5LF-8734].

[323] Anupam Chander, *The Racist Algorithm?*, 115 MICH. L. REV. 1023, 1040 (2017); Jessica Leber, *The Machine-Readable Workforce*, MIT TECH. REV. (May 27, 2013), www.technologyreview.com/s/514901/the-machine-readable-workforce/ [https://perma.cc/XB2D-WY85] (describing Xerox's approach as of 2013).

the software accordingly. The operator or owner does not have the ability to do that, thus the burden of liability should be spread between these two principals in light of their respective ability to oversee the behavior and actions of the AI agent-guard.

In the second example, we assume the tech company is the operator but not the owner of the algorithm. It is reasonable to assume this company had a say with regards to the metrics they hold dear in an attempt to optimize the algorithm's result for their purpose.[324] In this case, the tech company had the ability to monitor and guide the actions of the algorithms. The decisions of the algorithm by themselves are discriminatory, but the decision to implement those decisions were of the company and not of the other entities that may be considered to be principals.[325] It is true the tech company could not have predicted the discriminatory results of the algorithm given the metrics it chose, but if it decided to implement its recommendation, they should be viewed as the sole principal of the AI agent. Alternatively, if the algorithm was given to the company with a built-in metric system decided upon by the designer or programmer and implemented by the tech company, there would have been a better case for multiple principals and shared responsibility between the tech company and the creators of the algorithm.[326]

C. The Respondeat Superior Doctrine

Establishing the premise that an AI entity creates an agency relationship with its surroundings, this Article turns to discuss the implied connection between strict liability regime and agency law via the *respondeat superior* doctrine ("let the master answer"), that is, vicarious liability.[327]

This doctrine is based on the notion that if a principal entrusts subordinates, in our case AI agents, with "inherently risk-bearing activities, fairness requires the superior to assume responsibility for the conduct of its subordinates in carrying out their duties."[328] In practice,

---

[324] *Id.* (noting that Xerox valued improvement of tenure).

[325] *Id.* (stating that Xerox chose not to implement the discriminatory decisions).

[326] These two examples are oversimplified and are only meant to present the process of thought of the agent-principal application. Many other assumptions will change the outcomes in the analysis. They are to be taken as basic instances to the application of the agency analogy but they in no way diminish the complexity of identifying the appropriate agent.

[327] *See* Kristopher-Kent Harris, *Drones: Proposed Standards of Liability*, 35 SANTA CLARA HIGH TECH. L. J. 65, 74–81 (2018) (analyzing the *respondeat superior* doctrine with regards to drones); Dafni Lima, *Could AI Agents Be Held Criminally Liable: Artificial Intelligence and the Challenges for Criminal Law*, 69 S. C. L. REV. 677, 692 (2018) (analyzing the *respondeat superior* doctrine as applied to AI entities for criminal responsibility).

[328] BENNETT, *supra* note 287, at 161.

one can claim the principal is in a better position to endure the damage or acquire insurance against potential liability claims than the agent itself (i.e., the principal is the best *pressure point* to incentivize).[329] This can also motivate the principal to choose its agents more carefully and invest more time and efforts in their selection, training, and subsequent monitoring.[330] This doctrine is designed in part to ensure tort victims will not be undercompensated due to an insolvent agent.[331] This rationale is particularly important when we discuss AI agents because they are inherently insolvent.[332]

Applying strict liability in the form of *respondeat superior* ensures the principal will be better able to control and adjust the levels of activity of her AI agent. Influencing one's levels of activity is not plausible in a negligence regime, which puts most of its weight on the question of

---

[329] *See* BENNETT, *supra* note 287, at 44.

[330] BENNETT, *supra* note 287, at 44. Several examples of this idea have been noted by courts. In *Kerl v. Rasmussen* the court declared that:

Vicarious liability under the doctrine of respondeat superior depends upon the existence of a master/servant agency relationship. Vicarious liability under respondeat superior is a form of liability without fault -- the imposition of liability on an innocent party for the tortious conduct of another based upon the existence of a particularized agency relationship. As such, it is an exception to our fault-based liability system, and is imposed only where the principal has control or the right to control the physical conduct of the agent such that a master/servant relationship can be said to exist.

682 N.W.2d 328, 331 (Wis. 2004). Likewise, *Convit v. Wilson* explained that:

We begin with the legal principles that will guide our analysis. "Vicarious liability . . . is merely a legal concept used to transfer liability from an agent to a principal[,]" and includes the theory of respondeat superior as developed in agency law. Under that theory, the responsibility of an agent for his own legally careless action is imputed to the principal. Thus, we have said that "in cases involving derivative liability under the doctrine of respondeat superior, the master's liability is limited to that of the servant, the only active tortfeasor . . . ."

980 A.2d 1104, 1114 (D.C. 2009); *see also* Judah v. Reiner, 744 A.2d 1037, 1039–40 (D.C. 2000) (stating that to show liability of a principal, it must be shown that an agency relationship exists and the agent was acting within the scope of the agency relationship); Giles v. Shell Oil Corp., 487 A.2d 610, 611 (D.C. 1985) (stating that to succeed under a *respondeat superior* theory, the existence of a master-servant relationship must be shown and that the employee was acting in the scope of his employment when the harm occurred); Smith v. Jenkins, 452 A.2d 333, 335 (D.C. 1982) (stating that "[g]enerally an agency relationship results when one person authorizes another to act on his behalf subject to his control, and the other consents to do so"). Other decisions focus on consent and control as a key component of the agency relationship. *See, e.g.*, Burlington Indus., Inc. v. Ellerth, 524 U.S. 742 (1998); Troyer v. Nat'l Futures Ass'n, 290 F. Supp. 3d 874 (N.D. Ind. 2018); Tillman v. U.S. Dep't of Health & Human Services, No. 16-CV-1335-PP, 2016 U.S. Dist. LEXIS 178749 (E.D. Wis. Dec. 27, 2016); Henderson v. Charles E. Smith Management, Inc., 567 A.2d 59, 62 (D.C. 1989).

[331] Justin Sevier, *Vicarious Windfalls*, 102 IOWA L. REV. 651, 660 (2017).

[332] Wein, *supra* note 142, at 110 ("The obligation to make restitution falls on the master as the financially responsible individual, rather than on the servant as the morally responsible party."). However, in the AI context, it is difficult to view AI agents as "morally responsible" for their actions. *Id.*

whether the harmful action fulfilled the duty of care obligation.[333] Strict liability, on the other hand, enables the principals to expand their control on the scope of the activity.[334] This incentivizes them to better train and guide their AI agents in an attempt to reach the optimal level of activity via the internalization of the costs associated with these high levels of activity.[335]

To establish the *respondeat superior* doctrine, there should be an appropriate relationship between the superior and its subordinate, and an appropriate connection between that relationship and the conduct of the subordinate which leads to the damage.[336] In order to decide whether vicarious liability should apply in a specific case, the common test courts utilize examines whether the agent was acting "in the course of the employment," when the damage occurred.[337] This is meant to create a distinction between acts carried out by the agent for which the principal will not be held liable and those acts for which she will.[338] In the AI context, this distinction does not exist, as there are no acts that can be carried out by the AI agents that will exceed the liability scope of the principal given their mono-purpose and "excessive loyalty" traits.

The *Restatement (Third) of Agency* discusses a principal's liability for an agent operating within the "scope of employment."[339] It states that "[a]n employee acts within the scope of employment when preforming

---

[333] *See, e.g.*, Tom Baker, *On the Genealogy of Moral Hazard*, 75 TEX. L. REV. 237, 270 (1996).

[334] A major criticism of strict liability is that eventually everyone insures themselves and they will be covered for damages. Therefore, it may not mean anything. While it may be impossible, or at the very least difficult, to discuss strict liability without discussing the effects insurance law has on this liability regime, in the framework of this article, it cannot be done and will be reserved for a future project. *See, e.g.*, Anat Lior, *Insurability of Artificial Intelligence Algorithms and Robots – A Different Version of the Same Policy*, THE FEDERMANN CYBER SECURITY CENTER– CYBER LAW PROGRAM (Aug. 5, 2019), csrcl.huji.ac.il/blog/anat-lior-Insurability-AI). It is sufficient to say that moral hazards have yet to undermine strict liability as an appropriate and apt regime and insurance companies have found, and will find again, ways to incentivize policy holders to manage their risk in a cautious way, such as premiums, caps, and deductibles. *See id.*; Tom Baker, *Liability Insurance as Tort Regulation: Six Ways that Liability Insurance Shapes Tort Law in Action*, 12 CONN. INS. L.J. 1 (2005); Louis Kaplow, *An Economic Analysis of Legal Transitions*, 99 HARV. L. REV. 509 (1986); Daniel Keating, *Pension Insurance, Bankruptcy and Moral Hazard*, 1991 WIS. L. REV. 65 (1991); Steven L. Schwarcz, *Systemic Risk*, 97 GEO. L.J. 193 (2008).

[335] Steven Shavell, *Liability for Accidents*, HANDBOOK OF LAW AND ECONOMICS 139, 146–147 (2007); *see* Balkin, *supra* note 17, at 1240. Balkin refers to this with regards to the "levels of decision-making activity" an algorithm takes, and the internalization of the costs produced by these activities by the business which are operating them.

[336] Catholic Church Welfare Society v. Various Claimants, UKSC 56, 3 WLR 1319 [2012]; BENNETT, *supra* note 287, at 161.

[337] SHERMAN, P. TECUMSEH, WORKMEN'S COMPENSATION LAW: "PERSONAL INJURY BY ACCIDENT ARISING OUT OF AND IN THE COURSE OF THE EMPLOYMENT" 18 (1916).

[338] BENNETT, *supra* note 287, at 162.

[339] RESTATEMENT (THIRD) OF AGENCY § 7.07 (AM. LAW INST. 2006).

work assigned by the employer or engaging in a course of conduct subject to the employer's control."[340] It states further that "[a]n employee's act is not within the scope of employment when it occurs within an independent course of conduct not intended by the employee to serve any purpose of the employer."[341] An employer-employee relationship is a sub-category of an agent-principal relationship and are therefore applicable to other types of relationships within the agency framework.[342] In the AI context, the mono-purpose characteristic of AI agents renders the applicability of the end of this clause in the AI context void. AI agents lack the ability to act in "an independent course of conduct not intended to further any purpose of the employer" given the control its human principal has over its assigned tasks and the lack of a truly autonomous aspect to its existence.[343]

In his article, Bathaee also supports the application of the *respondeat superior* doctrine with regards to autonomous AI entities in light of the black-box problem.[344] However, he limits this application to certain circumstances, "[w]hen the AI operates autonomously in a mission-critical setting or one that has a high possibility of externalizing the risk of failure on others."[345] However, this conclusion that vicarious liability will be less appropriate in "less dangerous or mission-critical settings" is incorrect for two main reasons.[346] First, the differentiation between these two scenarios is not clear and the possibility of externalizing the risk of failure on others is high in both situations given the black-box issue. Second, the risk of chilling "a large swath of desirable AI applications" is always relevant when AI entities are involved and is not necessarily more justified in the second set of circumstances Bathaee describes in light of the first reason mentioned above.[347] Furthermore, the chilling effect problem (i.e., stifling innovation) is a major concern in the process of regulating AI liability.[348] However, advocating for no-liability or lowering the liability bar (in the form of negligence supervision of the principal) will lead to problematic

[340] *Id.*
[341] *Id.*
[342] *See generally* Charles Davant IV, *Employer Liability for Employee Fraud: Apparent Authority or Respondeat Superior*, 47 S.D. L. REV. 554 (2002) (explaining *respondeat superior* and apparent authority through the employer-employee relationship).
[343] RESTATEMENT (THIRD) OF AGENCY § 7.07 cmt. b (AM. LAW INST. 2006); *see supra* Part IV, Section C.
[344] Yavar Bathaee, *The Artificial Intelligence Black Box and the Failure of Intent and Causation*, 31 HARV. J.L. & TECH. 889, 934 (2018).
[345] *Id.* at 935.
[346] *Id.*
[347] *Id.*
[348] *Id.* at 928

results in the AI industry that will eventually prevent it from internalizing its inflicted damages and improving its practices.[349]

A *terminology* note must be made prior to ending this sub-section about the primary liable party in this vicarious liability context. In a relationship of a human principal and an AI agent, the latter cannot be found liable, so technically the principal is not *vicariously* liable but rather *primarily* liable.[350] The distinction between the two does not matter to the third-party that was injured, but it does matter to the principal itself when she desires to recoup the damages she paid to the injured party.[351]

This distinction has practical ramifications. For example, a principal can purchase an insurance policy to protect herself in cases of vicarious liability but not in cases of "primary liability engendered by acts of an agent attributed to the principal."[352] When we discuss an AI agent, which lacks the ability to assume responsibility over its actions, the only entity we can claim as responsible is the human principal or principals pulling its strings. Thus, concepts of primary and vicarious liability should be treated differently in the AI agent context than in the case of a human agent. To prove vicarious liability, there is no obligation to point to an entity which is primarily liable, especially in the AI context where we know the AI lacks the capability to be held liable. The human principal or principals will be named as liable and in fact they will be held *primarily* liable for the actions of their AI agents.

V. CONCLUSION

AI entities today, and more so in the future, are an essential part of modern society.[353] They are causing damages and are bound to cause even more as their integration into our homes and industries increase.[354] Using legal analogies to reason the appropriate liability regime that should apply on AI entities is a common method for handling new phenomena in general, and new technologies in particular.[355] However, choosing the appropriate legal analogy is not a stand-alone decision. It is influenced by our regulatory purposes and the way we comprehend AI entities as part of our society.

---

[349] *See supra* Part IV, Section B.

[350] *See* Paulius Cerka, Jurgita Grigiene & Gintare Sirbikyte, *Liability for Damages Caused by Artificial Intelligence*, 31 COMPUTER L. & SECURITY REV. 376 (2015) (explaining that the likelihood of damages caused by AI is real).

[351] BENNETT, *supra* note 287, at 168.

[352] *Id.*

[353] Michelle Evans, *Artificial Intelligence is Expected to be the Most Impactful Technology on Commerce*, FORBES (Jan. 16, 2019), www.forbes.com/sites/michelleevans1/2019/01/16/artificial-intelligence-is-expected-to-be-the-most-impactful-technology-on-commerce/#51ce92dd3b23 [https://perma.cc/J4KK-DGGT].

[354] *See* Cerka et al., *supra* note 350, at 892–93.

[355] Graham, *supra* note 65, at 1242.

Treating AI entities as AI agents, which are under the control and guidance of human principals, is the most accurate analogy we can use to represent their relationship with our society. This is so given their mono-purpose feature, the instrumental value they serve in accomplishing tasks humans assign to them, and the fact all other non-agency legal analogies can be reduced to a three-way agency relationship (because AI entities are in essence AI judgment-proof agents). Moreover, even if one does not agree with this legal analogy, the strict liability regime that stands behind it, in the form of *respondeat superior*, is the most appropriate in the AI liability context.

When discussing reflective equilibrium, Rawls stated that "[i]f the scheme as a whole seems on reflection to clarify and to order our thoughts, and if it tends to reduce disagreements and to bring divergent convictions more in line, then it has done all that one may reasonably ask."[356] AI entities as AI agents is the theory that accomplishes this reasonable request and grants us clarity in a very cloudy and unpredictable field. It is not problem-free, however. Identifying the appropriate principal or principals is a problem that will need to be resolved within our judiciary, insurance, and administrative systems, but it is a manageable and productive problem nonetheless and makes it the most appropriate theory to follow.

Agents, including AI agents, come in all shapes and sizes and their utilization under the assumption that their main goal is to provide a service to a human being is intuitive. This may very well change once we acknowledge (or realize) they have outgrown that purpose and by doing so are no longer agents, let alone our agents. This conclusion may also change in the future if we decide AI entities have reached an intelligence level which entitles them to new rights and obligations. These two scenarios will lead to the localization of the agency analogy—once the technology behind an AI entity changes, the analogy used to describe them is threatened and the rational that stands behind it is undermined.[357] This does not necessarily mean the analogy is moot or erased, but rather it is localized, meaning, it may be applicable with regards to a specific area within the field, but not with the subject matter as a whole.[358] This is echoed in Rawls's reflective equilibrium process which leads to the localization of some analogies due to the back and forth reasoning process that changes as AI advances. The localization of

[356] See RAWLS, *supra* note 7, at 53.

[357] Blavin & Cohen, *supra* note 48, at 285.

[358] For example, treating the internet as the information superhighway may still be appropriate with regards to emails, but not the internet as a whole. Blavin & Cohen, *supra* note 48, at 269–74.

the agency analogy will then force us to embark once again on our back and forth reasoning voyage.[359]

If indeed that day comes,[360] and the regulatory purposes that stands at the basis of the justification for treating AI as agents will be no more, a new or old legal analogy should be considered. If AI entities will cease to be judgment-proof agents and will become full transparent agents, the regulatory structure of agency relationship will be relevant no more and other non-agency legal analogies, including personhood, could be back in the race to lead the appropriate liability regime.

But until that day, AI judgment-proof agents will continue to possess the potential to wreak havoc without the ability of their potential victims to protect themselves. Holding their human principals strictly liable for the AI agents' mischiefs will incentives them to create a safer environment in today's *algorithmic society*,[361] one that is much needed now, and even more so in the future.

---

[359] The reflective equilibrium process teaches us that as AI advances and presents more independent will, older metaphors may become localized. This view is in line with Blavin and Cohen's conclusion that metaphors' future application may become irrelevant. Blavin & Cohen, *supra* note 48, at 269–74. For example, treating robots as property or slaves might be deemed as appropriate for dependable robots of the previous decade, such as Innvo. Labs robotic dinosaur Pleo and the first generation of Sony's dog, AIBO, neither of which possessed the capability to make decisions on their own. *See* Geoffrey Fowler, *AIBO the Dog will Melt your Heart with Mechanical Precision*, THE WASHINGTON POST (Sept. 19, 2018), www.washingtonpost.com/technology/2018/09/18/aibo-robot-dog-will-melt-your-heart-with-mechanical-precision/?noredirect=on&utm_term=.21a6785258c4 [https://perma.cc/59EK-BSLT] (noting that in 2018, Sony released a new and improved version of AIBO. As of the time of its release, it costs $2,900). Pleo is an animatronic pet dinosaur toy. For more examples of these types of "basic" robots, see Darling, *supra* note 97, at 215. Still, this metaphor is problematic these days given the growing proficiencies possessed by AI entities, which may lead to a day when we will no longer set in advance the goals and behavior of AI entities by code alone. Thus, general principles change over time due to the impact of specific cases and the adoption of new and better-suited analogies.

[360] *See supra* Part III, Section B.

[361] Balkin, *supra* note 17, at 1219.