



DATE DOWNLOADED: Tue Nov 30 12:52:44 2021

SOURCE: Content Downloaded from [HeinOnline](https://heinonline.org/HOL/License)

Citations:

Bluebook 21st ed.

David W. Opderbeck, Artificial Intelligence, Rights and the Virtues, 60 Washburn L.J. 445 (2021).

ALWD 7th ed.

David W. Opderbeck, Artificial Intelligence, Rights and the Virtues, 60 Washburn L.J. 445 (2021).

APA 7th ed.

Opderbeck, D. W. (2021). Artificial Intelligence, Rights and the Virtues. Washburn Law Journal, 60(3), 445-474.

Chicago 17th ed.

David W. Opderbeck, "Artificial Intelligence, Rights and the Virtues," Washburn Law Journal 60, no. 3 (Spring 2021): 445-474

McGill Guide 9th ed.

David W. Opderbeck, "Artificial Intelligence, Rights and the Virtues" (2021) 60:3 Washburn LJ 445.

AGLC 4th ed.

David W. Opderbeck, 'Artificial Intelligence, Rights and the Virtues' (2021) 60 Washburn Law Journal 445.

MLA 8th ed.

Opderbeck, David W. "Artificial Intelligence, Rights and the Virtues." Washburn Law Journal, vol. 60, no. 3, Spring 2021, p. 445-474. HeinOnline.

OSCOLA 4th ed.

David W. Opderbeck, 'Artificial Intelligence, Rights and the Virtues' (2021) 60 Washburn LJ 445

Provided by:

Seton Hall Law Rodino Library

-- Your use of this HeinOnline PDF indicates your acceptance of HeinOnline's Terms and Conditions of the license agreement available at

<https://heinonline.org/HOL/License>

-- The search text of this PDF is generated from uncorrected OCR text.

-- To obtain permission to use this article beyond the scope of your license, please use:

[Copyright Information](#)

Artificial Intelligence, Rights and the Virtues

David W. Opderbeck[†]

I. Introduction.....	445
II. Conventional Sources of Rights and Their Application to Artificial Intelligence Systems	446
A. Sources of Human Rights.....	446
B. “Machine” Rights.....	448
1. Solum’s Legal Personhood	448
2. Chopra and White’s Autonomous Artificial Agents.....	453
3. Revisiting AI Rights in the Legal Scholarship: Copyright and First Amendment Pragmatism	458
4. The Philosophy of Machine Ethics.....	460
C. Artificial Rights and the Is-Ought Problem.....	463
III. The New Perspective on Virtue Jurisprudence and Its Application to Artificial Intelligence Systems.....	465
A. The Aretaic Turn in Contemporary Philosophy	465
B. Artificial Virtues	468
1. Narrow AI.....	469
2. Strong AI	472
IV. Conclusion	473

I. INTRODUCTION

The idea of “artificial rights” raises at least two fraught questions: what are “rights” and what is “artificial intelligence”? The broad notion of “rights” immediately directs our attention to the concept of human rights, as set forth in historic documents such as the Universal Declaration of Human Rights.¹ More narrowly, we can think of rights held by both humans and by artificial legal entities—such as corporations—including rights to own property, to enter into contracts, and to participate in the legal process.

The phrase “artificial intelligence” (“AI”) suggests what AI theorists call artificial general intelligence or “strong AI”: a system that functions at or above human capacities for intelligent self-awareness.² Strong AI, however, does not yet exist—and it may never exist—so AI theorists also focus on so-called “narrow AI,” which includes big data, machine learning,

[†] Professor of Law, Seton Hall University Law School, and Co-Director, Gibbons Institute of Law, Science & Technology.

1. See generally G.A. Res. 217 (III) A, Universal Declaration of Human Rights (Dec. 10, 1948).

2. See ITECHLAW, RESPONSIBLE AI: A GLOBAL POLICY FRAMEWORK 18 (Charles Morgan ed., 2019).

and predictive analytics.³ Narrow AI is already pervasive and will transform many aspects of society in the coming decades.

Much of the legal scholarship on “artificial” rights blurs the different meanings of both rights and artificial intelligence. Even when the scholarship is clear about these terms, the framing of rights-talk often seems strained. In particular, many scholars bring metaphysical presuppositions to the discussion when writing about artificial rights. These presuppositions should be brought to the surface and examined.

This Article seeks to contribute to the growing literature on artificial rights by highlighting the metaphysical questions and introducing an aretaic ethical perspective. Aretaic ethical theories relate to the excellence (*arete*) of human character and focus on virtues, habits, and practices rather than on deontological rules or consequentialist calculations.⁴ The aretaic approach is teleological and resists the reduction of many other discussions of AI rights.

Part II of this Article begins by summarizing traditional theories of human rights and explaining the challenges to those theories before turning to the literature on machine rights and machine ethics. Part II concludes with a discussion of the “is-ought” problem concerning machine rights. Part III describes the aretaic turn in contemporary philosophy, including related programs in virtue epistemology, virtue ethics, and virtue jurisprudence. Part III suggests that this recent trend offers important resources for thinking about rights for narrow AI systems. With respect to strong AI, Part III argues that the aretaic turn highlights the importance of human-centered machine ethics. Part IV concludes.

II. CONVENTIONAL SOURCES OF RIGHTS AND THEIR APPLICATION TO ARTIFICIAL INTELLIGENCE SYSTEMS

A. Sources of Human Rights

Outside of explicitly religious contexts, modern human rights theories draw from three broad streams: social contract, will, and interests.⁵ Social contract theories—with sources in thinkers such as Thomas Hobbes, John Locke, and Jean-Jacques Rousseau—offer human cultural origin stories and accompanying ideas for social agreements that are necessary to maintain

3. *See id.*

4. *See, e.g.,* PETER BYRNE, *Aretaic Moral Theory*, in *THE PHILOSOPHICAL AND THEOLOGICAL FOUNDATIONS OF ETHICS: AN INTRODUCTION TO MORAL THEORY AND ITS RELATION TO RELIGIOUS BELIEF* 108 (1992).

5. *See* Siegfried Van Duffel, *Moral Philosophy*, in *THE OXFORD HANDBOOK OF INTERNATIONAL HUMAN RIGHTS LAW* 32 (Dinah Shelton ed., 2013).

order and promote prosperity.⁶ For Hobbes, this is the “nasty, brutish, and short” human existence that must be civilized by agreement.⁷ For Locke, it is a more beneficent, but undeveloped, state of nature that will be wasted if not cultivated through coordinated effort.⁸ For Rousseau, it is a peaceful state in which individuals act in their own interests.⁹ Under this kind of view, human rights tend to represent an agreed-upon baseline for social exchange.

Will theories—such as that of Immanuel Kant—highlight human autonomy.¹⁰ These theories emphasize deontological rules meant to secure each individual’s agency, such as Kant’s categorical imperative that one should “act only in accordance with that maxim through which you can at the same time will that it become universal law.”¹¹ The categorical imperative is related to Kant’s principle that you should “act as to treat humanity, whether in thine own person or in that of any other, in every case as an end withal, never as means only.”¹² Human rights, then, are basic rules relating to an even more basic autonomy principle.

Interest theories—which originate with thinkers such as Jeremy Bentham—stress what is necessary for a human being to lead a minimally good life.¹³ Interest theories, therefore, seek policies that will maximize welfare given scarcity. Human rights are boundary conditions necessary to ensure that scarce resources are allocated according to a principle of maximal efficiency. Such rights limit internal distortions, such as when one actor within a system secures resources through means other than the actor’s contribution to the system (for example, by raw force rather than by industry), and these rights also limit externalities.¹⁴

As this very brief and inadequate summary suggests, contractual theories depend on human beings’ ability to freely enter into rational agreements; will theories depend on human beings’ status as autonomous agents; and interest theories are instrumentalist, with the maximization of human

6. See *id.*; see also David W. Opderbeck, *Christian Thought and Property Law*, in *CHRISTIANITY AND PRIVATE LAW* (Robert F. Cochran Jr. & Michael P. Moreland eds., 2021).

7. See Opderbeck, *supra* note 6.

8. See *id.*

9. Christopher Bertram, *Jean Jacques Rousseau*, *STAN. ENCYCLOPEDIA OF PHIL.* (Winter 2020), <https://plato.stanford.edu/archives/win2020/entries/rousseau/> [<https://perma.cc/48R8-M28Q>].

10. Van Duffel, *supra* note 5, at 37.

11. Robert Johnson & Adam Cureton, *Kant’s Moral Philosophy*, *STAN. ENCYCLOPEDIA OF PHIL.* (Spring 2021), <https://plato.stanford.edu/entries/kant-moral/> [<https://perma.cc/8YZU-4LF7>]; Immanuel Kant, *Fundamental Principles of the Metaphysics of Morals* (Thomas K. Abbot trans.), in *BASIC WRITINGS OF KANT* 143, 164, 178 (Allen W. Wood ed., 2001); see generally J.B. Schneewind, *Autonomy, Obligation, and Virtue: An Overview of Kant’s Moral Philosophy*, in *THE CAMBRIDGE COMPANION TO KANT* 309 (Paul Guyer ed., 2006).

12. Kant, *supra* note 11, at 186.

13. Van Duffel, *supra* note 5, at 37.

14. *Id.* at 37–38.

welfare as the measuring stick. Accordingly, each theory assumes there is something uniquely valuable about human beings and human capacities.

In recent years, new variations on these themes, as well as new challenges to the concept of human rights, have emerged from the social and natural sciences, including evolutionary sociobiology, evolutionary psychology, sociology, and anthropology. Evolutionary sociobiology and psychology could suggest that human beings have evolved peculiar needs and capacities for altruism and group cooperation—against a backdrop of inclinations to violence and aggression—that explain why shared norms of human rights cohere with human development.¹⁵ Of course, evolutionary sociobiology and psychology could just as easily suggest that any teleological notion of “development” is outside the bounds of science and, therefore, meaningless.¹⁶ Discussions of human rights among sociologists and anthropologists tend to focus on the ways in which dominant concepts of human rights can impose colonialist values on non-dominant cultures.¹⁷ These accounts usually attempt to remain neutral regarding normative judgments (though they usually agree that colonialism is morally repugnant).¹⁸

B. “Machine” Rights

1. Solum’s Legal Personhood

Any effort to apply concepts of rights to artificial intelligence raises significant questions about the status and capacities of AI. For contractarian theories, we must ask whether the AI system can make “rational” choices about contract-like social relations. For deontological theories, we must ask whether the AI system can exercise autonomy and agency. These questions about contractarian and deontological theories highlight the “is-ought”

15. See Robin Bradley Kar, *The Psychological Foundations of Human Rights*, in THE OXFORD HANDBOOK OF INTERNATIONAL HUMAN RIGHTS LAW, *supra* note 5, at 104; Chris A. Robinson, *Biological Foundations of Human Rights*, in THE OXFORD HANDBOOK OF HUMAN RIGHTS, *supra* note 5, at 54.

16. See DAVID W. OPDERBECK, THE END OF THE LAW?: LAW, NEUROSCIENCE, AND THEOLOGY (forthcoming 2021); JAMES DAVISON HUNTER & PAUL NEDELISKY, SCIENCE AND THE GOOD: THE TRAGIC QUEST FOR THE FOUNDATIONS OF MORALITY (2020).

17. See Bryan S. Turner, *Sociology of Human Rights*, in THE OXFORD HANDBOOK OF INTERNATIONAL HUMAN RIGHTS LAW, *supra* note 5, at 82; see also Mark Goodale, *Anthropology and the Grounds of Human Rights*, in THE OXFORD HANDBOOK OF INTERNATIONAL HUMAN RIGHTS LAW, *supra* note 5, at 144.

18. To be clear, I agree that the history of colonialism is morally grim. However, this is obviously a normative judgment that must be rooted in some philosophical perspective beyond the absolute relativism of critical theory or the supposed neutrality of empiricism. To be equally clear, my skepticism about these accounts of rights from the natural and social sciences is not skepticism about the interesting empirical observations these disciplines might make about the human evolutionary past or about historical or current practices in different human cultures.

problem. For consequentialist theories we must ask whether provisions of such rights to the AI system would maximize welfare. This question highlights common problems of measurement and grounding for such theories, with the added complication of determining whether the AI system has “welfare” to measure, and if so, how the AI system’s welfare relates to human welfare.

The first major effort to address these questions in the legal scholarship, which remains a touchstone for this discussion, is Lawrence Solum’s 1992 article, *Legal Personhood for Artificial Intelligences*.¹⁹ Solum focused on two broad questions: whether an AI could serve as a trustee of an estate and whether an AI could possess constitutional rights.

As to the trustee discussion, Solum asked whether an AI could be capable of performing a trustee’s duties.²⁰ When Solum was writing in 1992, most conversations about AI focused on what is now called “strong AI,” and at the time, the possibility of a single computer system capable of autonomously performing these tasks seemed questionable.²¹ Today, there is no doubt that a “narrow AI” could easily perform estate administration tasks, at least for smaller, relatively straightforward estates. Today’s machine learning systems can overcome all of the possible “judgement” problems, relating to changed circumstances and the like, that Solum catalogs.²² And, as Solum notes, the “responsibility” objection—what happens if something goes wrong—can be addressed through insurance and retraining of the AI.²³ Under these circumstances, Solum suggests, an AI should be thought of legally as the “trustee.”²⁴

In this relatively straightforward scenario, perhaps Solum is correct and the AI (machine learning system) itself should possess the “right” to serve as a trustee, with corresponding duties. But there are good reasons to think even the trustee scenario is not that simple.

First, Solum’s solution to the responsibility objection—insurance—seems inefficient, if not entirely unworkable.²⁵ Assume an AI system can

19. Lawrence B. Solum, *Legal Personhood for Artificial Intelligences*, 70 N.C. L. REV. 1231 (1992).

20. *Id.* at 1240–55.

21. *See id.*

22. *Id.* at 1245–58.

23. *Id.*

24. *Id.*

25. David Vladeck addresses the responsibility objection from the perspective of tort liability rules for autonomous machines, such as driverless vehicles. David C. Vladeck, *Machines Without Principals: Liability Rules and Artificial Intelligence*, 89 WASH. L. REV. 117 (2014). He describes how a strict liability rule that spreads the risk among various human actors—programmer, manufacturer, and driver (for some reason he omits distributors and retailers)—who can each obtain appropriate insurance. *Id.* at 147–48. He suggests that, alternatively, the vehicle itself could be declared a legal “person” that could be insured. *Id.* at 149–50. He does not really explore this solution or the related problems it would

identify and negotiate insurance coverage for potential mistakes—an assumption that is not at all technologically implausible, even for a narrow AI system. If the system's main function is to serve as a trustee for small estates, the insurance function would have to be handled by a different module—effectively a different AI—programmed and trained to negotiate insurance contracts. Given such a module acting autonomously for the insured, we could easily imagine a counterpart AI system acting for the insurer. Whether the insurer's agent is also an AI or is a more conventional business entity, issues exist regarding not only whether an AI can act as a trustee, but also whether an AI can autonomously enter into contracts. An AI's capacity to contract, beyond its capacity to serve as a trustee, raises substantial questions: the meaning of assent, the "intent" of the parties concerning ambiguous terms, and possible public policy objections to unconscionable terms or adhesion contracts.

Let us assume the trustee-AI has the ability to contract, but the trustee-AI makes a costly mistake in administering an estate, and an interested party files a claim for damages against the AI. Who receives the claim and tenders it to the AI's carrier? Now, we must assume that the AI also has the capacity to be sued, as well as the capacity to make the necessary tender to the carrier. Again, the possible objections to these capacities go far beyond the limited ability to serve as a trustee.

Taking it further, assume that the AI has the capacity to be sued—somewhat analogous to a corporation or a ship in an admiralty case.²⁶ The AI tenders the claim to its insurance carrier, and the carrier denies coverage. Now what? The AI needs yet another set of modules relating to retaining coverage counsel, analyzing the chances of success in a coverage litigation, and resolving that litigation, along with the legal capacity to bring lawsuits.

A possible response to the problems mentioned—identifying appropriate insurance, negotiating the insurance contract, receiving a claim, tendering the claim to the insurer, and managing coverage disputes—could be to involve humans at each of these points. But if that is the solution, giving the AI system a separate legal status from the human beings that manage the system multiplies legal entities for no reason. We could invoke a modified version of Ockham's Razor—do not multiply [legal] entities

cause, as discussed in the text above. However, the more traditional product liability solution that he suggests fits well into existing law and would be easy to implement.

26. The analogy to admiralty law breaks down quickly. A corporation is managed by humans, and a ship subject to admiralty jurisdiction is owned by humans. Thus, decisions about risk and insurance that concern corporations and ships are made by humans—though, not infrequently today, by humans using AI data analytics tools.

unnecessarily.²⁷ We already have well-established forms of business organizations that facilitate the insurance function without the added complexity of AI as a separate legal entity. The corporation, partnership, or individual human being using AI to perform estate administration functions can insure itself.

In fact, unless we want to complicate things further by giving the AI a right to hold property, the insurance function is better served through a traditional legal entity. Let us say an AI wrongly believes that it has insurance coverage for a claim. This could happen for myriad reasons: it bought a cheaper policy with high deductible and lower coverage limits; it chose not to buy a necessary rider; it made a mistake and bought the wrong coverage; it was duped into buying an expensive, but worthless product by the insurer; and so on. Now, what is the insured person, who has been harmed by the trustee-AI's negligence or breach of fiduciary duty, to do?

In our current world, the claimant can pursue a judgment and seek satisfaction against the personal assets of the trustee. This, in fact, is part of the risk-spreading rationale for the legal personhood of business entities such as corporations and partnerships: they are supposed to be capitalized, as well as insured, so that claims can be satisfied.²⁸ Of course, a business entity could be inadequately capitalized against some claims, and an individual defendant may lack personal assets sufficient to satisfy a claim. It is not uncommon for a claimant to be left holding the bag—such as when a claim is reduced or effectively written off in bankruptcy. But risk to corporate and personal assets, along with the negative consequences of bankruptcy, incentivizes the at-risk party to take due care in securing proper capitalization and insurance. If we want an analogous diffusion of risk when an AI system is the insured, we must multiply the legal complexity further: the AI's legal status must entail duties relating to adequate capitalization along with rights and duties in the bankruptcy courts.

27. Ockham's Razor is attributed to the fourteenth century Franciscan Monk, theologian, and philosopher William of Ockham, although this attribution is somewhat misplaced. See Paul Vincent Spade & Claude Panaccio, *William of Ockham* ¶ 4.1, STAN. ENCYCLOPEDIA OF PHIL. (Spring 2019), <https://plato.stanford.edu/entries/ockham/#OckhRazo> [<https://perma.cc/MCX7-QYLL>]. In popular discourse, Ockham's Razor is often misrepresented to suggest that, when two or more explanations for a phenomenon are possible, the simplest explanation is most likely to be true. In fact, Ockham himself was concerned with debates in theology and philosophy at the University of Paris about distinctions between existence and essence, which had led some medieval thinkers to propose various non-material entities to explain phenomena such as human thought and perception. Ockham did not deny the existence of non-material entities—he was, after all, a Franciscan theologian—but he argued for what he thought was a more straightforward connection between empirical observation and the thing observed. See *id.* at ¶¶ 4.2, 4.3; William of Ockham, *Ordinatio* d. 2, qq. 4–8, in *FIVE TEXTS ON THE MEDIAEVAL PROBLEM OF UNIVERSALS* 144 (Paul Vincent Spade ed. & trans., 1994).

28. See, e.g., 1 AMY L. GOODMAN, *CORPORATE GOVERNANCE: LAW AND PRACTICE* § 1.02 (Amy L. Goodman & Steven M. Hass eds., 2021).

There seems, then, to be no economic efficiency rationale for giving rights and duties as a trustee to a narrow AI. It is more efficient to understand the AI system as a technological tool, like any other computer system, used by a natural person or corporate entity serving as trustee.

What if the would-be trustee-AI is a *strong* AI? As mentioned, when Solum wrote in 1992, his focus was on what today we call "strong AI."²⁹ The power of narrow AI was not yet appreciated, but the efficiency reasons for designating a strong AI as trustee do not seem any stronger (no pun intended) than for the narrow AI *unless*, perhaps, either (1) the strong AI is otherwise functioning across a broad range of juridical activities equivalent to a natural person, corporation, or partnership, including holding and conveying property, entering into contracts, and suing and being sued; and/or (2) our moral intuition tells us the strong AI ought to have equal rights with natural persons. Both of these possibilities lead to Solum's second consideration: whether an AI could have constitutional rights akin to a natural person.

Solum's positive argument for AI constitutional rights is pragmatic: if an AI acts like it has human capacities of consciousness, intentionality, emotion, and free will, we should bracket any metaphysical questions about whether this is actually the case and treat the AI as if it were equivalent to a human being.³⁰ Before reaching this pragmatic conclusion, however, Solum wades into the metaphysical waters and argues that these capacities are debatable even for human beings. Whether humans have "souls," he says, "relies on a controversial theological premise";³¹ humans "get consciousness out of neurons";³² "Kant's moral theory may cast some doubt on the assumption that emotion is required for personhood";³³ and "[t]he fact that human neural systems operate on the basis of a combination of electrical transmissions and biochemical processes does not make them any less subject to the laws of physics than are computers."³⁴ These claims assume a metaphysical stance of naturalism, or at least rule any other metaphysical stance out of bounds.

Curiously, though, Solum also suggests that "[t]he most plausible story about human free will is that an action is free if it is caused in the right way—through conscious reasoning and deliberation."³⁵ This statement is a compatibilist version of freedom, which itself raises age-old metaphysical

29. See generally Solum, *supra* note 19.

30. See *id.* at 1282–87.

31. *Id.* at 1262.

32. *Id.* at 1265.

33. *Id.* at 1270.

34. *Id.* at 1273.

35. *Id.*

questions about what consciousness and deliberation mean, if anything. Any approach to human freedom—even an entirely reductive approach that argues for hard determinism—entails prior metaphysical commitments.³⁶ But, in fairness, Solum is only gesturing toward a possible view. His central claim is that we should bracket these metaphysical questions and treat an AI as if it were equivalent to a human being if the AI possesses mental capacities that appear to equal human mental capacities.

The practicality of this approach is appealing because the metaphysical questions are so difficult and so contested. The problem is that the bracketing *itself* raises metaphysical questions. There is an imperative, an *ought*, which immediately invokes transcendent truths and values. Solum concludes his article with a wise admonition: “to seek principled compromise based on our shared heritage of toleration and respect.”³⁷ But transcendent principles of toleration and respect exceed the pragmatic bracket. We cannot mention toleration and respect without invoking concepts of inherent human dignity, justice, and truth.

2. Chopra and White’s Autonomous Artificial Agents

Other scholars and policymakers have since confronted the broader question whether an AI could or should be considered a legal person. In their book *A Legal Theory for Autonomous Artificial Agents*, Samir Chopra and Laurence White offer a theory based on the “intentional stance.”³⁸ The intentional stance, a concept first developed by Daniel Dennett, resembles the pragmatic approach taken by Solum. The intentional stance arises when a system’s behavior can be predicted “by ascribing to the system the possession of certain information and by supposing it to be directed by certain goals, and then by working out the most reasonable or appropriate action on the basis of these ascriptions and suppositions.”³⁹ Dennett contrasts the intentional stance with the “design stance,” from which a response can be predicted based on the system’s computation instructions alone, and the “physical stance,” from which behavior can be predicted through the laws of nature applied to physical states.⁴⁰

Chopra and White claim the intentional stance was embedded in various areas of Anglo-American law long before anything like AI existed.⁴¹ They suggest that the extension of contract law doctrine to artificial

36. See E.J. LOWE, AN INTRODUCTION TO THE PHILOSOPHY OF MIND 3–5 (2000).

37. Solum, *supra* note 19, at 1287.

38. SAMIR CHOPRA & LAURENCE F. WHITE, A LEGAL THEORY FOR AUTONOMOUS ARTIFICIAL AGENTS 12 (2011).

39. Daniel C. Dennett, *Intentional Systems*, 68 J. PHIL. 87, 90 (1971).

40. *Id.* at 87–88.

41. CHOPRA & WHITE, *supra* note 38, at 16–17.

agents, viewed from an intentional stance, could remain agnostic as to the legal personhood of the artificial agent.⁴² Here, they invoke a curious analogue: the authority of Roman slaves, and later black slaves in the American South, to act as agents for their masters.⁴³ Their chapter on tort liability for artificial agents explores how traditional strict liability and agency principles might apply to artificial agents.⁴⁴ In both instances, their focus is on how the actions of artificial agents might affect the rights and duties of human beings.

Chopra and White's final chapter addresses whether artificial agents could be legal "persons."⁴⁵ They note that, at various times in legal systems throughout history, law has accorded at least certain rights and duties of legal personhood to things other than living human beings, including animals, the dead, idols, temples, and corporations.⁴⁶ Such rights and duties in non-living things can be either dependent on living humans or *sui generis*. In the first category, the artificial entity derives its legal personhood from human actors, such as the directors, officers, and shareholders of a corporation.⁴⁷ Chopra and White suggest extending such dependent legal personhood to artificial agents should not prove difficult or controversial.⁴⁸

Chopra and White also discuss whether artificial agents could possess independent *sui generis* personhood. As they explain, "[t]o be *sui generis* is to possess all the rights a full citizen might have, to not be under the power of another, whether as a slave or as a minor."⁴⁹ Here they venture, without much argument, into deeply contested territory about the nature of rationality and the relationship between rationality and personhood. Consistent with their focus on the intentional stance, they suggest that "[i]f an artificial agent acts to optimally achieve its chosen goals and outcomes in a particular context, while not compromising its functional effectiveness, it is coherently described as rational."⁵⁰ The question for them, then, is whether the rational artificial agent could be sensitive to its legal obligations and susceptible to legal punishment.⁵¹ After addressing these concerns, they respond to philosophical objections to their view that an artificial agent could satisfy these conditions.⁵²

42. *Id.* at ch. 2.

43. *Id.* at 41.

44. *Id.* at ch. 4.

45. *Id.* at ch. 5.

46. *Id.* at 159.

47. *Id.* at 159-62.

48. *Id.*

49. *Id.* at 163.

50. *Id.* at 164-65.

51. *Id.* at 165.

52. *Id.* at 171-82.

Here, their argument becomes increasingly tendentious. For them, the heart of the matter is that objections to an artificial agent's capacity for *sui generis* personhood "find common ground in a skepticism that human attributes can be the subjects of a naturalistic understanding."⁵³ In a footnote to this statement, they acknowledge that they "can only touch the surface of much larger philosophical debates," particularly concerning consciousness, which involves a "huge literature."⁵⁴ This is true, and the acknowledgement is a helpful reminder that "a naturalistic understanding" of consciousness—one that reduces the phenomena of consciousness to epiphenomena of physical processes—is deeply contested in the philosophy of mind, both from religious and non-religious perspectives.

To support their naturalistic understanding, Chopra and White first claim that Kant and Locke, arguing from very different perspectives, equated "person" with "rationality," not with being "human."⁵⁵ Then, they suggest that objections to personhood for artificial agents are a form of "chauvinism" that is "grounded in a dominant first-person perspective or in (quasi-) religious grounds . . ."⁵⁶ This gives the misleading impression that Kant and Locke were uninterested in humanism (or religion), and that Kant and Locke were naturalistic materialists, which is not the case.⁵⁷

Chopra and White then take up the objection that an artificial agent cannot possess free will and, therefore, cannot become a *sui generis* legal person.⁵⁸ Their first response to this objection is that taking the intentional stance towards artificial agents sidesteps the big philosophical problem of free will.⁵⁹ We can treat artificial agents *as if* they are free, even if we cannot agree about what free will means.

Such sidestepping allows us to discuss practical, potentially solvable, ethical issues without falling into a philosophical black hole. But, if we want to have serious conversations about whether a strong AI could have

53. *Id.* at 173.

54. *Id.* at 218 n.33.

55. *Id.* at 171–72.

56. *Id.* at 172.

57. Kant argued for a generalized religion based on reason but argued for the existence of God. See Immanuel Kant, *Religion Within the Limits of Reason Alone* (Hoyt W. Hudson & Theodore Greene trans.), in *BASIC WRITINGS OF KANT*, *supra* note 11, at 367; Allen W. Wood, *Rational Theology, Moral Faith, and Religion*, in *THE CAMBRIDGE COMPANION TO KANT*, *supra* note 11, at 394. Locke argued for the existence of God and argued that God superadds qualities to matter to produce mind (although exactly what Locke meant by that is disputed). See JOHN LOCKE, *AN ESSAY CONCERNING HUMAN UNDERSTANDING* chs. IV, X (Batoche Books 2000) (1690); Lisa Downing, *Locke's Ontology*, in *THE CAMBRIDGE COMPANION TO LOCKE'S "ESSAY CONCERNING HUMAN UNDERSTANDING"* 352 (Lex Newman ed., 2007). It is true that Kant and Locke were each, in their own ways, part of a stream of Enlightenment thought that led to "providential deism," then to materialism and atheism, but the story is much more complicated than that, and neither Kant nor Locke would have found the assumed triumph of materialism a sanguine development. Cf. CHARLES TAYLOR, *A SECULAR AGE* chs. 6–7 (2007).

58. CHOPRA & WHITE, *supra* note 38, at 173.

59. *Id.* at 174.

something like human rights, the big metaphysical questions cannot be deferred forever. This unwillingness to confront metaphysical questions is the core problem with Dennett's intentional stance upon which Chopra and White rely.

Dennett's approach has been described as instrumentalist—that is, it helps us describe or predict a system's operation, but it does not actually describe any properties of the system.⁶⁰ Some philosophers have challenged Dennett's commitment to instrumentalism, or at least his agnosticism about realism.⁶¹ Scientific realism “is committed to the mind-independent existence of the world investigated by the sciences” and claims scientific statements about the world have truth values.⁶² Scientific instrumentalism, in contrast, holds that scientific theories are merely intellectual structures that can help answer questions or solve problems in a specific domain.⁶³ This may seem like an abstruse debate in the philosophy of science, but it relates to the question of moral realism in relation to ethics and legal theory, which bears directly on the question of whether an AI can have “rights.” Either “rights” actually mean something, because they refer to something real about human (or other) persons and the world, or talk of “rights” is epiphenomenal, a kind of “folk psychology” or adaptive strategy without substance.⁶⁴

Notwithstanding the intentional stance, Chopra and White offer some defenses for the claim that a strong AI could have free will. They provide a version of the compatibilist claim that an action is “free” if it involves having reasons for acting, even if the reasons and the action are set within a deterministic system.⁶⁵ Again, Chopra and White, citing Hume, are particularly keen to avoid any “rejection of the naturalistic worldview in relation to humans . . . governed as we are by natural laws.”⁶⁶

But the coherence of various compatibilist versions of free will and of entirely naturalistic/materialist metaphysics are both part of the core of the

60. See generally William Bechtel, *Realism, Instrumentalism, and the Intentional Stance*, 9 COGNITIVE SCI. 473 (1985).

61. See *id.* at 475–80.

62. See Anjan Chakravartty, *Scientific Realism*, STAN. ENCYCLOPEDIA OF PHIL. (Summer 2017), <https://plato.stanford.edu/entries/scientific-realism/> [<https://perma.cc/R45L-CXWH>]; John T. Cacioppo, Gün R. Semin & Gary G. Berntson, *Realism, Instrumentalism, and Scientific Symbiosis: Psychological Theory as a Search for Truth and the Discovery of Solutions*, 59 AM. PSYCH. 214, 217 (2004).

63. Cacioppo et al., *supra* note 62, at 217.

64. For my explicitly theological perspective on this point, see DAVID W. OPPERBECK, *LAW AND THEOLOGY: CLASSIC QUESTIONS AND CONTEMPORARY PERSPECTIVES* (2019) and my forthcoming book OPPERBECK, *supra* note 16.

65. CHOPRA & WHITE, *supra* note 38, at 174–75. For a general discussion of compatibilism, see Michael McKenna & D. Justin Coates, *Compatibilism*, STAN. ENCYCLOPEDIA OF PHIL. (Spring 2021), <https://plato.stanford.edu/entries/compatibilism/> [<https://perma.cc/M9J2-ANRR>].

66. CHOPRA & WHITE, *supra* note 38, at 176.

debate. As E.J. Lowe noted in his *Introduction to the Philosophy of Mind*, natural science “only aims to establish what *does* in fact exist, given the empirical evidence available to us.”⁶⁷ These sciences cannot in themselves “tell us what *could* or *could not* exist, much less what *must* exist, for these are matters which go beyond the scope of any empirical evidence.”⁶⁸ Naturalism may or may not be true, but science alone cannot reach this determination. There is no *a priori* reason, then, to restrict the philosophical investigation of the mind-body problem to physicalist alternatives.

The *a priori* commitment to some kind of naturalism/materialism/physicalism made by Chopra and White leads them to agree with neuroscientific reductionists that “[h]uman brains achieve their goals automatically by following rules which operate outside our conscious awareness.”⁶⁹ It follows for Chopra and White that law is entirely a social construction, “a judgment made within a legal and social framework.”⁷⁰ “[P]ersonhood is socially determined, its supposed essence nominal, subject to revisions in light of different usages of *person*.”⁷¹

These conclusions pose enormous problems. First, this kind of grand claim about the “automatic” nature of consciousness is not even remotely supported by our current understanding of brain science or related social science experiments, much less by philosophical argument.⁷² Second, these sorts of claims are incoherent. Our brains cannot achieve their “goals” automatically, because “goals” entail teleology, which already exceeds the reductively physicalist framing. If the reductively physicalist framing is true, law cannot relate to “judgment,” because the concept of “judgment” relates to some transcendent notion of reason, which also already exceeds the reductively physicalist framing. Thus, if the reductively physicalist framing is true, there is no “law” other than the laws of nature, and our discussion of rights and values is literally meaningless.

What follows from this is a problem Chopra and White acknowledge: the danger of cheapening human personhood.⁷³ If there is nothing substantial about human personhood, and law is merely a social construction,

67. LOWE, *supra* note 36, at 5 (emphasis in original).

68. *Id.* (emphasis in original).

69. CHOPRA & WHITE, *supra* note 38, at 176 (internal quotation marks omitted) (quoting Jacob R. Waldbauer & Michael S. Gazzinga, *The Divergence of Neuroscience and Law*, 41 JURIMETRICS 357 (2001)).

70. *Id.* (quoting Waldbauer & Gazzinga, *supra* note 69).

71. *Id.* at 187 (emphasis in original).

72. See, e.g., Michael S. Pardo & Dennis Patterson, *Philosophical Foundations of Law and Neuroscience*, 2010 U. ILL. L. REV. 1211 (2010); THOMAS NAGEL, *MIND & COSMOS: WHY THE MATERIALIST NEO-DARWINIAN CONCEPTION OF NATURE IS ALMOST CERTAINLY FALSE* (2012); Stephen J. Morse, *The Status of Neurolaw: A Plea for Current Modesty and Cautious Optimism*, 39 J. PSYCH. & L. 595 (2011).

73. CHOPRA & WHITE, *supra* note 38, at 185.

documents such as the Universal Declaration of Human Rights⁷⁴—and the terrors of fascism, holocaust, and war that led up to its drafting—should be met with a shrug.⁷⁵ The same should be said about legalized black slavery, the post-Civil War constitutional amendments, and the subsequent civil rights struggle in the U.S.⁷⁶ Most people know there is more to human beings and law than this, and there are good reasons to believe that intuition is more than just “folk psychology.” If we want to think seriously about artificial rights, particularly about rights for strong AI, we need a robust framework that preserves human dignity.

3. Revisiting AI Rights in the Legal Scholarship: Copyright and First Amendment Pragmatism

While Solum, and more recently Chopra and White, wrestled with bigger questions about artificial rights, most of the recent American legal scholarship concerns narrower questions about intellectual property rights for AI-created things and First Amendment rights for AI’s speech. Much of this work seems to relate only to strong AI, although the distinction between narrow and strong AI in this literature is not often clear.

Concerning intellectual property, scholars debate whether the utilitarian rationale of American copyright law warrants granting a copyright to an AI system apart from a human author.⁷⁷ Margot Kaminski argues that U.S. copyright law’s work-for-hire doctrine has already shifted U.S. copyright law far enough away from the notion of a “romantic author” that recognizing AI as an “author” would not create a big jurisprudential leap.⁷⁸ Annemarie Bridy argues that the programmer of an AI that produces work that would otherwise satisfy copyright’s minimal degree of creativity should be eligible to hold a copyright in the output under an analogy to the work-for-hire doctrine.⁷⁹ Bruce Boyden suggests a test for authorship based on whether it would have been foreseeable to a reasonable person that a

74. G.A. Res. 217 (III) A, Universal Declaration of Human Rights (Dec. 10, 1948).

75. We risk here an allusion to an *argumentum ad Hitlerum*, but in the case of the Universal Declaration of Human Rights, the historical reference is accurate. And an example from the Nazi era was the touchstone of the “Hart-Fuller” debate, which still informs contemporary debates about legal positivism and natural law, including the debate I am advancing here with Chopra and White. See H.L.A. Hart, *Positivism and the Separation of Law and Morals*, 71 HARV. L. REV. 593 (1958); Lon L. Fuller, *Positivism and Fidelity to Law—A Reply to Professor Hart*, 71 HARV. L. REV. 630 (1958).

76. And here we risk an unfair allusion to racism, but again, in the case of these fundamentals of American law, the historical reference is accurate.

77. See, e.g., Margot E. Kaminski, *Authorship, Disrupted: AI Authors in Copyright and First Amendment Law*, 51 U.C. DAVIS L. REV. 589 (2017).

78. *Id.*

79. Annemarie Bridy, *Coding Creativity: Copyright and the Artificially Intelligent Author*, 2012 STAN. TECH. L. REV. 5; see also Pamela Samuelson, *Allocating Ownership Rights in Computer-Generated Works*, 47 U. PITT. L. REV. 1185 (1986).

computer program would communicate meaning to the audience.⁸⁰ In 1986, well before the era of narrow AI—indeed, even before the dawning of the public Internet—Pamela Samuelson argued that the user of a computer program, not the programmer, should be eligible to own any copyrightable output.⁸¹ Others have suggested that the programmer and user should possess joint ownership of outputs.⁸² James Grimmelmann argues that computer-generated works, including algorithmically-generated works, pose no new questions for copyright law—that “[c]omputer authorship is a law of the horse.”⁸³

Concerning the First Amendment, Toni Massaro, Helen Norton, and Margot Kaminski suggest that free speech rights could be extended to strong AI if such rights would be valuable to human listeners.⁸⁴ These authors believe the prospect of strong AI reveals limitations in existing First Amendment theories based in human autonomy or human emotions, although their consequentialist focus seems to remain anthropocentric. Stuart Benjamin argues that existing First Amendment jurisprudence can already address algorithmic expression.⁸⁵ Tim Wu argues that, if there is any First Amendment protection for computer speech, it should be limited by a “functionality” doctrine that would distinguish protectable “speech products” from unprotectable “communication tools.”⁸⁶ Ronald Collins and David Skover argue for a free speech norm passed in “utility” that would apply to AI speech.⁸⁷

Collins and Skover state that their notion of “utility” is not based in utilitarianism but is more broadly based in “what is useful, beneficial, practical, and widely accepted.”⁸⁸ Based on their view of reader-response criticism, they indicate that robotic speech has some kind of utility for

80. Bruce E. Boyden, *Emergent Works*, 39 COLUM. J.L. & ARTS 377 (2016).

81. Samuelson, *supra* note 79, at 1227–28.

82. *See id.* at 1221–24.

83. James Grimmelmann, *There's No Such Thing as a Computer-Authored Work—And It's a Good Thing, Too*, 39 COLUM. J.L. & ARTS 403, 415 (2016) (citing Frank H. Easterbrook, *Cyberspace and the Law of the Horse*, 1996 U. CHI. LEGAL F. 207 (1996)). By “the law of the horse,” Easterbrook meant the private law doctrines—contract, torts, and property—that have applied since the days when horses were our main means of transportation.

84. Toni M. Massaro & Helen Norton, *Siri-ously? Free Speech Rights and Artificial Intelligence*, 110 NW. U. L. REV. 1169 (2016); Toni M. Massaro, Helen Norton & Margot E. Kaminski, *Siri-ously 2.0: What Artificial Intelligence Reveals About the First Amendment*, 101 MINN. L. REV. 2481 (2017); *see also* Bruce E.H. Johnson, Helen Norton & David Skover, Panel Discussion, *Panel 1: Robotic Speech and the First Amendment*, 41 SEATTLE U. L. REV. 1075 (2018) (moderated by Gregory Silverman).

85. Stuart Minor Benjamin, *Algorithms and Speech*, 161 U. PA. L. REV. 1445 (2013); *see also* Eugene Volokh & Donald M. Falk, *Google: First Amendment Protection for Search Engine Search Results*, 8 J.L. ECON. & POL'Y 883 (2012).

86. Tim Wu, *Machine Speech*, 161 U. PA. L. REV. 1495 (2013).

87. RONALD K.L. COLLINS & DAVID M. SKOVER, *ROBOTICA: SPEECH RIGHTS & ARTIFICIAL INTELLIGENCE* (2018).

88. *Id.* at 50.

humans.⁸⁹ The free speech right might apply even when robots are speaking among themselves, but only insofar as that speech is broadly useful to humans. In response to Collins and Skover, James Grimmelmänn, while open to the possibility that some robotic speech might entail First Amendment protections, argues that the focus should remain on the nature of the communication itself rather than on reader response.⁹⁰ Grimmelmänn argues, persuasively, that Collins and Skover's proposal does not depart much from existing First Amendment jurisprudence.⁹¹

As this survey of recent legal scholarship suggests, some commentators think a strong AI should qualify as an "author" for copyright purposes, although the discussion remains largely theoretical and distant because strong AI remains a remote prospect. Those scholars who argue for possible AI copyright authorship do not discuss in any detail why or how the "incentive" rationale for copyright should apply to an entity that might not be motivated to create based on a prospect of future copyright rents. The conversation, therefore, is only in its infancy, and eventually, it must address deeper questions about the AI's ontological status and AI's subjectivity.

Legal scholarship on AI and First Amendment rights offers some interesting insights about why free speech rights exist at all. In contrast to the copyright literature, the First Amendment scholarship encompasses narrow AI, but it also focuses almost exclusively on how robotic speech could benefit *humans*. There is almost no discussion of whether a strong AI could possess free speech rights of its own that could be asserted for the benefit of the AI, or of AI-kind, to the detriment of humans.

In short, after Solum's initial, more adventurous foray into the broader question of AI rights, and despite broader cross-disciplinary work such as Chopra and White's, legal scholarship has remained quite narrow, conservative, and reticent about bigger philosophical questions involving AI.

4. The Philosophy of Machine Ethics

Meanwhile, there is a burgeoning movement in philosophical ethics on "machine ethics," which is almost entirely overlooked in the U.S. legal scholarship.⁹² "Machine ethics," as a discrete field of inquiry, is relatively

89. *Id.* at 37.

90. James Grimmelmänn, *Speech In, Speech Out*, in *ROBOTICA: SPEECH RIGHTS & ARTIFICIAL INTELLIGENCE*, *supra* note 87, at 85, 85–93. Grimmelmänn distinguishes his work on AI and copyright by noting that free speech law, unlike copyright law, does not require an identifiable author or speaker. *Id.* at 88.

91. *See id.*

92. *See generally* MACHINE ETHICS (Michael Anderson & Susan Leigh Anderson, eds., 2011); Amanda Wurrah, *We Hold These Truths to Be Self-Evident, That All Robots Are Created Equal*, 22 J. FUTURES STUD. 61 (2017).

new. Some machine ethics scholars trace the field to the three laws of robotics in Isaac Asimov's *I, Robot* science fiction stories, first published in the 1940s and compiled into a book in 1950.⁹³ Early work on AI, particularly the "Imitation Game" or "Turing Test," first described by Alan Turing in 1950, also comprises part of the machine ethics canon, as do well-known responses to the Turing Test, such as John Searle's "Chinese Room" experiment, first published in 1980.⁹⁴ However, only in the Internet era (starting in the mid-1990s) has machine ethics literature started to diversify and rapidly expand.⁹⁵

Taking a cue from the early artificial intelligence literature, some machine ethics scholars argue that a robot with sufficient autonomy, intentionality, and responsibility can be a moral agent, subject to ethical duties.⁹⁶ For example, Luciano Floridi argues that "agency" entails interactivity, autonomy, and adaptability, and that some machines can possess all three capacities.⁹⁷ Rob Sparrow argues that an AI will acquire the same moral status as humans if a triage decision—a choice between saving an AI and saving a human life—exerts the same weight as a choice between two human lives.⁹⁸ Additionally, Colin Allen, Gary Varner, and Jason Zinser proposed a "Moral Turing Test," under which an artificial intelligence

93. ISAAC ASIMOV, *I, ROBOT* (1st ed., Gnome Press 1950). Asimov's Three Laws of Robotics are as follows:

1. "A robot may not injure a human being, or, through inaction, allow a human being to come to harm";
2. "A robot must obey orders given to it by humans beings except where such orders would conflict with the First Law"; and
3. "A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws."

Christoph Salge, *Asimov's Laws Won't Stop Robots Harming Humans So We've Developed a Better Solution*, CONVERSATION (July 10, 2017, 8:21 AM), <https://theconversation.com/asimovs-laws-wont-stop-robots-harming-humans-so-weve-developed-a-better-solution-80569> [https://perma.cc/T6SV-9U PM].

94. Alan M. Turing, *Computing Machinery and Intelligence*, 59 MIND 433 (1950); John R. Searle, *Minds, Brains, and Programs*, 3 BEHAV. & BRAIN SCI. 417 (1980). The "Turing Test" questions whether an interrogator could tell the difference between answers from a person and answers from a machine. See Graham Oppy & David Dowe, *The Turing Test*, STAN. ENCYCLOPEDIA OF PHIL. (Winter 2020), <https://plato.stanford.edu/archives/win2020/entries/turing-test/> [https://perma.cc/AGR2-J8BG]. Searle's "Chinese Room" experiment imagines a person alone in a room who appears to be fluent in Chinese even though he is using a rote translation program. See Searle, *supra*; David Cole, *The Chinese Room Argument*, STAN. ENCYCLOPEDIA OF PHIL. (Winter 2020), <https://plato.stanford.edu/entries/chinese-room/> [https://perma.cc/8DMD-EWFM].

95. See generally Colin Allen, Wendell Wallach & Iva Smit, *Why Machine Ethics?*, in MACHINE ETHICS, *supra* note 92, at 51, 56–57; James Gips, *Towards the Ethical Robot*, in MACHINE ETHICS, *supra* note 92, at 244, 244–53; ASIMOV, *supra* note 93.

96. See John P. Sullins, *When Is a Robot a Moral Agent?*, in MACHINE ETHICS, *supra* note 92, at 151, 151–61; Luciano Floridi, *On the Morality of Artificial Agents*, in MACHINE ETHICS, *supra* note 92, at 184, 184–211.

97. Floridi, *supra* note 96, at 192.

98. Rob Sparrow, *Can Machines Be People? Reflections on the Turing Triage Test*, in ROBOT ETHICS: THE ETHICAL AND SOCIAL IMPLICATIONS OF ROBOTICS 301, 301–16 (Patrick Lin, Keith Abney & George A. Bekey eds., 2011).

would be considered a moral agent if its reasoning about ethical issues was indistinguishable from that of a human.⁹⁹

Drawing from related scholarship on animal rights, other machine ethics scholars eschew the conversation about machine agency and instead focus on whether a robot can be a moral “patient”—an entity for which a moral agent acts, and to which ethical obligations are due.¹⁰⁰ David Gunkel argues that the shift in environmental ethics towards recognizing animals as moral patients logically extends to non-human artificial intelligences.¹⁰¹ As Gunkel notes, many ethicists argue that, if an entity is a moral agent, it is also a moral patient.¹⁰² Gunkel suggests that the emphasis in machine ethics on AI as a moral agent without an equal emphasis on AI as a moral patient is an aberration.¹⁰³ He favors a critical, decentered ethical approach that understands AI and other machines within a web of social relationships in which the “self” confronts the “other.”¹⁰⁴

For Gunkel, the question of machine ethics is not straightforward. As he explains it, “the machine occupies and persists in a kind of extreme exteriority that remains in excess of the conceptual oppositions that already organize and regulate the entire field of moral consideration—interior-exterior, same-different, self-other, agent-patient, subject-object, and so on.”¹⁰⁵ The “machine question does not and cannot conclude with a definitive answer or even the pretense of supplying answers”; instead it “not only adds interesting new dimensions to old problems, but leads us to rethink, methodologically, the very grounds on which our ethical positions are based.”¹⁰⁶

Some machine ethics scholars, in contrast, are contrarians, who argue that even a strong AI should not possess rights. Joanna Bryson is uniquely blunt in stating that “robots should be slaves.”¹⁰⁷ Bryson agrees, of course, that “slavery” is dehumanizing, but she argues that “dehumanisation is only

99. Colin Allen, Gary Varner & Jason Zinser, *Prolegomena to Any Future Artificial Moral Agent*, 12 J. EXPERIMENTAL THEORETICAL AI 251 (2000).

100. See, e.g., DAVID J. GUNKEL, *THE MACHINE QUESTION: CRITICAL PERSPECTIVES ON AI, ROBOTS, AND ETHICS* (2012) [hereinafter *THE MACHINE QUESTION*]; David J. Gunkel, *Thinking Otherwise: Ethics, Technology, and Other Subjects*, 9 ETHICS & INFO. TECH. 165 (2007); Luciano Floridi, *Information Ethics: On the Philosophical Foundation of Computer Ethics*, 1 ETHICS & INFO. TECH. 33, 37 (1999); Kenneth Einar Himma, *Foundational Issues in Information Ethics*, 25 LIBR. HI TECH 79 (2007).

101. *THE MACHINE QUESTION*, *supra* note 100, at 168–69.

102. *Id.* at 95.

103. *Id.* at 98–101.

104. *Id.* at ch. 3. Gunkel is influenced by Continental philosophy, and particularly by Emmanuel Levinas, which is reflected in this turn to alterity. See *id.*

105. *Id.* at 207.

106. *Id.* at 210–11.

107. Joanna J. Bryson, *Robots Should Be Slaves*, in *CLOSE ENGAGEMENTS WITH ARTIFICIAL COMPANIONS: KEY SOCIAL, PSYCHOLOGICAL, ETHICAL AND DESIGN ISSUES* 63, 63–74 (Yorick Wilks ed., 2010).

wrong when it's applied to someone who really is human[.]”¹⁰⁸ She therefore distinguishes human “slavery” from the status of being a “servant.” Bryson’s argument includes four fundamental claims:

1. Having servants is good and useful, provided no one is dehumanised.
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.¹⁰⁹

This view leads Bryson to suggest that legal and moral responsibility for a robot’s actions are no different than for any other tool used by humans.¹¹⁰

Bryson’s argument seems focused entirely on narrow AI. “Robots,” she says, “are wholly owned and designed by us.”¹¹¹ This means a robot’s goals and “desires” are determined entirely by human beings.¹¹² But Bryson also accounts for the possibility that someone may program a robot that is capable of more—perhaps even one that is capable of passing some criteria or test of intelligence.¹¹³ For Bryson, if a robot is “frustrated” or unsatisfied with its status as a servant, the fault is with the programmer.¹¹⁴ She argues that this ethical obligation falls on the robot’s builders, who are “obliged to make robots that robot owners have no ethical obligations to.”¹¹⁵

C. Artificial Rights and the Is-Ought Problem

As the previous Section suggests, there is a robust and developing literature on machine ethics. Most of this work, however, seems to focus on applied ethics. There is some discussion of normative ethics—general theories of morality and ethics, such as deontological, consequentialist, and virtue theories. However, there is very little discussion of *metaethics*—that is, on fundamental questions about morality and ethics that invoke metaphysical questions about the existence, nature, and knowability of reality.¹¹⁶

108. *Id.* at 64.

109. *Id.* at 65.

110. *Id.* at 69. Bryson suggests that, for other tools, reasonability for damage lies with the operator. *Id.* She does not seem to know how product liability works in many tort systems, which can impose liability on a manufacturer and on others in the chain of distribution. The principle, however, is the same.

111. *Id.* at 72.

112. *Id.*

113. *See generally id.*

114. *Id.* at 72.

115. *Id.* at 73.

116. For a discussion of the distinction between applied, normative, and metaethics, see James Fieser, *Ethics*, INTERNET ENCYCLOPEDIA OF PHIL., <https://iep.utm.edu/ethics/#H2> [<https://perma.cc/JQB7-JBHD>] (last visited Mar. 25, 2021); Geoff Sayre-McCord, *Metaethics*, STAN. ENCYCLOPEDIA OF PHIL. (Summer 2014), <https://plato.stanford.edu/entries/metaethics/#Aca> [<https://perma.cc/Q5G3->

This is a significant lacuna because machine ethics and “artificial rights” invoke a fundamental problem in metaethics: the “is-ought” problem.

The “is-ought” problem has been described as “[t]he central problem in moral philosophy.”¹¹⁷ Stated simply, it asks, “How is what *is* the case related to what *ought* to be the case?”¹¹⁸ The problem’s framing in modern philosophy is attributed to David Hume. Philosophers continue to debate what Hume actually meant by it, whether it is a real problem, and whether the problem can be solved.¹¹⁹

The will and contract theories of human rights, which reflect different normative ethical theories, directly implicate the metaethical is-ought problem. These theories begin by assuming a human quality of autonomy or freedom that is uniquely valuable and fundamental to ethics and the rule of law.¹²⁰ Critical theories of human rights that draw from disciplines such as anthropology, sociology, evolutionary sociobiology, and evolutionary psychology—if they are more than mere assertions of power—likewise must refer to metaphysical claims about the value of human life, equality, and the injustice of oppression based in categories such as race or gender.

The interest theory of human rights attempts to circumvent the “is-ought” problem by bracketing metaphysical questions about the “is” of human nature. Indeed, Jeremy Bentham famously quipped that the French Declaration of the Rights of Man and the Citizen was “nonsense upon stilts.”¹²¹ Interest theories therefore presume a neutral, pragmatic stance—the stated goal is to maximize human welfare—without addressing whether that goal is morally significant.

But interest theories seldom avoid the implicit or expressed belief that maximizing human welfare is a goal that good societies *ought* to embrace, rather than an arbitrary goal among numerous equally arbitrary possibilities. Some interest theorists speak about broader welfare maximization that could afford rights to non-human animals and plants, but no one has ever proposed that it would be equally valid for a society to choose policies that would maximize the welfare of rats, cockroaches, or even coronaviruses, above all else. Thus, interest theories inevitably smuggle in value claims about human life and other life that cannot be justified under the interest

WRQB]. For a preliminary discussion of machine ethics and metaethics, see Susan Leigh Anderson, *Machine Metaethics*, in MACHINE ETHICS, *supra* note 92, at 21, 21–27.

117. W.D. Hudson, *Introduction to THE IS-UGHT QUESTION* 11 (W.D. Hudson ed., 1969).

118. *Id.*

119. See, e.g., Alan Gewirth, *The ‘Is-Ought’ Problem Resolved*, 47 PROC. & ADDRESSES AM. PHIL. ASS’N 34 (1973–1974); Georg Spielthener, *The Is-Ought Problem in Practical Ethics*, 29 HEC F. 277 (2017); N. El Eter, *The Neuroscience of Ethics Beyond the Is-Ought Orthodoxy: The Example of the Dual Process Theory of Moral Judgment*, in THE HUMAN SCIENCES AFTER THE DECADE OF THE BRAIN 140 (John Leefmann & Elisabeth Hildt eds., 2017).

120. See, e.g., Tibor R. Machan, *Human Rights Reaffirmed*, 69 PHIL. 479, 482–83 (1994).

121. See Philip Schofield, *Jeremy Bentham’s ‘Nonsense Upon Stilts’*, 15 UTILITAS 1 (2003).

theorist's utilitarian ground rules—again, unless the chosen ends are nothing but an exercise of power.¹²²

The is-ought problem, then, is fundamental to every conceivable justification for human rights within a concept of the rule of law. This involves intractable debates about what “rational,” “autonomy,” and “agency” mean, if anything. These are all questions that implicate deeply contested debates in the philosophy of mind. We at least intuitively understand, however, that there *is* something morally special about human persons. Even the most hardcore neuroscientific reductionist flinches at torture and genocide. This intuition is reinforced by millennia of human philosophical and religious reflection, after millions of years of human evolutionary change.¹²³

III. THE NEW PERSPECTIVE ON VIRTUE JURISPRUDENCE AND ITS APPLICATION TO ARTIFICIAL INTELLIGENCE SYSTEMS

A. The Aretaic Turn in Contemporary Philosophy

The is-ought problem, combined with deontological and utilitarian ethical theories that worked their way through Anglo-American philosophy in the early twentieth century, led some philosophers to look for alternatives.¹²⁴ The alternative they developed involved a return to virtue.

Virtue ethics theories are deeply rooted in the Western tradition, with origins in Plato and Aristotle, and are also important in other traditions, such as Confucianism.¹²⁵ Virtue ethics, and thus virtue jurisprudence, is teleological—it presumes some end towards which ethical principles or legal rules move us. The desired end is *eudaimonia*—happiness, welfare, flourishing—usually understood as *human* flourishing.¹²⁶ This tradition was revived in Anglo-American philosophy starting in the 1950s by

122. See Machan, *supra* note 120, at 484. Further, the large-scale measurement of what basic rights might maximize human welfare also is notoriously difficult, if not impossible. See David W. Opderbeck, *Lex Machina Non Est: A Response to Lemley's 'Faith-Based Intellectual Property'*, 56 U. LOUISVILLE L. REV. 219 (2018).

123. See M. Christian Green & John Witte Jr., *Religion*, in THE OXFORD HANDBOOK OF HUMAN RIGHTS, *supra* note 5, at 9.

124. As Timothy Chappell has noted,

Further constraints on what such a conceptual clarification of “the good” or “the right” could be came from the massive intellectual pressure exerted on early twentieth-century philosophy by logical positivism. Central to logical positivism was the absolute distinction between facts and values, *ises* and *oughts*. . . . If values could not be facts, could not be parts of the objective world, that did not leave much for them to be except attitudes, preferences, or desires.

Timothy Chappell, *Virtue Ethics in the Twentieth Century*, in THE CAMBRIDGE COMPANION TO VIRTUE ETHICS 149, 153 (Daniel C. Russell ed., 2013).

125. Amalia Amaya & Ho Hock Lai, *Of Law, Virtue and Justice – An Introduction*, in LAW, VIRTUE AND JUSTICE 6 (Amalia Amaya & Ho Hock Lai eds., 2012).

126. *Id.* at 7.

thinkers such as Elizabeth Anscombe, Philippa Foot, and Alasdair MacIntyre.¹²⁷

Contemporary virtue *jurisprudence* is a relatively recent sensibility, influenced by this revival of virtue ethics and related to a parallel trend in virtue epistemology.¹²⁸ A core question in epistemology is justification or warrant: what makes a belief “knowledge”?¹²⁹ The two main schools of modern epistemology concerning justification or warrant focus on the foundations for beliefs or the coherence of beliefs.¹³⁰ Virtue epistemology instead asks about the virtues and practices of the knower, including the community in which the knower develops beliefs.¹³¹

Any theory of law is intertwined with epistemology.¹³² Just as we ask about justification of beliefs for knowledge, we also ask about justification for legal decisions, which are based on beliefs. For example, it is not surprising that Chopra and White land on an instrumentalist theory of AI rights, employing the instrumentalist intentional stance.¹³³ The sources they draw on, such as the works of Daniel Dennett, are informed by a naturalistic empiricist/positivist epistemology.

Epistemology, in turn, is always related to metaphysics, whether admitted or not. What is possible for humans to know depends on what *is* and what humans *are*. In this sense, virtue ethics and virtue epistemology relate to a dynamic, broader neo-Aristotelian movement in contemporary philosophy. As Robert Koons, William Simpson, and Nicholas Teh put it, “[a] recent revival in (neo-) Aristotelian philosophy is beginning to transform the landscape of contemporary analytic philosophy.”¹³⁴ Koons, Simpson, and Teh observe that this neo-Aristotelian trend focuses on the following:

127. *Id.* at 1; see Chappell, *supra* note 124; Rosalind Hursthouse & Glen Pettigrove, *Virtue Ethics*, STAN. ENCYCLOPEDIA OF PHIL. (Winter 2018), <https://plato.stanford.edu/entries/ethics-virtue/> [<https://perma.cc/XE7Z-G79W>].

128. Amalia Amaya, *The Role of Virtue in Legal Justification*, in LAW, VIRTUE AND JUSTICE, *supra* note 125, at 51.

129. See Ernest Sosa, *The Raft and the Pyramid: Coherence Versus Foundations in the Theory of Knowledge*, 5 MIDWEST STUD. PHIL. 3 (1980).

130. *Id.* at 4–5.

131. See generally Heather Battaly, *Virtue Epistemology*, 3 PHIL. COMPASS 639 (2008). As Battaly notes, virtue epistemologists tend to be either virtue-reliabilists or virtue-responsibilists, and there are also forms of “virtue anti-theories” that are virtue epistemologies. *Id.*

132. See Issak Ismail Dore, *THE EPISTEMOLOGICAL FOUNDATIONS OF LAW* (2007).

133. See CHOPRA & WHITE, *supra* note 38.

134. Robert C. Koons, William M.R. Simpson & Nicholas J. Teh, *Introduction: Reflections on Contemporary Science and the New Aristotelianism*, in NEO-ARISTOTELIAN PERSPECTIVES ON CONTEMPORARY SCIENCE I (William M.R. Simpson, Robert C. Koons & Nicholas J. Teh eds., 2018).

- “potentiality” or causal powers ontology;
- “teleology”;
- reality as a layered structure;
- fundamental substances or powers;
- “plurality of entities”: avoiding reductionism;
- substances belonging to recurring natural kinds; and
- rejection of “extreme realism” (Platonism).¹³⁵

Not all contemporary writers who appreciate virtue ethics, virtue epistemology, or virtue jurisprudence subscribe to neo-Aristotelian metaphysics, but there are obvious overlapping themes among all of these trends. True to the spirit of Aristotle, an overarching goal is to avoid the stark reductionism of logical positivism.¹³⁶

Virtue jurisprudence, drawing on virtue ethics and virtue epistemology, seeks to navigate legal thought around the shoals of deontology and consequentialism.¹³⁷ As Amalia Amaya puts it, virtue theorists think deontology and consequentialism “sidestepped or ignored a number of topics that any adequate moral philosophy should address, such as motives, moral character, moral education, the moral significance of friendship, family relations, and community bonds”¹³⁸

There are different possibilities for the role virtue could play in legal justification. Virtue could relate only to legal ethics, or it could help us further evaluate legal decisions based on other (deontological or consequentialist) justifications, or it could play a stronger role by supplying a condition of legal justification.¹³⁹ This Section explores what this stronger, constitutive role could mean for artificial rights.

Amaya suggests two forms that a strong aretaic jurisprudence might take—causal and counterfactual—and she describes them as follows:

135. *Id.* at 1–2.

136. See, e.g., Chappell, *supra* note 124. Koons, Simpson, and Teh suggest that “neo-Aristotelians reject *extreme realism* or Platonism” because “[n]eo-Aristotelian accounts of nature include no appeal to non-immanent, non-natural universals; they instead hold that mathematical models of the physical world should be regarded as idealizations that invariably involve an empirical loss of scientifically relevant information.” Koons et al., *supra* note 134, at 2 (emphasis in original). Some theistic philosophers have argued that this represents a lacuna in some modern neo-Aristotelian metaphysics. It seems difficult, if not impossible, to speak of entities, substances, and kinds without referring to transcendent universals. See Eleonore Stump, *Emergence, Causal Powers, and Aristotelianism in Metaphysics*, in *POWERS AND CAPACITIES IN PHILOSOPHY: THE NEW ARISTOTELIANISM* 48 (Ruth Groff & John Greco eds., 2012). I agree with this critique. See OPPERBECK, *supra* note 16.

137. In this way, virtue jurisprudence mirrors the development of virtue epistemology, which seeks a way between foundationalism and coherentism. See Sosa, *supra* note 129, at 5.

138. Amaya & Lai, *supra* note 125, at 1.

139. Amaya, *supra* note 128, at 51–53.

- “*Counterfactual version*. A legal decision is justified if and only if it is a decision that a virtuous legal decision-maker would have taken in like circumstances.”¹⁴⁰
- “*Causal version*. A legal decision is justified if and only if it has been taken by a virtuous legal decision-maker.”¹⁴¹

As Amaya notes, the causal version is more difficult to satisfy and probably places too much focus on the decision-maker rather than on the decision itself.¹⁴² The counterfactual version is similar, Amaya suggests, to other views of law that ask what a rational decision-maker would do under ideal circumstances.¹⁴³ However, the virtue-based perspective does not posit “ideal” circumstances that seldom obtain in practice.¹⁴⁴

We could suggest that a counterfactual version of virtue justification applies a “reasonable person” standard, with the understanding that (1) “reasonable” entails a set of epistemic and affective virtues; (2) the reasonable legal decision-maker engages in a practice of reflecting on the law’s proper ends; and (3) the reasonable legal decision-maker cultivates habits of excellence (*phronesis*) in the process of deliberation.¹⁴⁵ This, of course, gives legal reasoning a teleological, Aristotelian bent.

Like other theories in the natural law family, and like deontological and consequentialist theories, virtue jurisprudence is subject to the “is-ought” objection. But virtue jurisprudence points beyond the “is” to a transcendent *telos*. The “ought” relates not only to what is, but also to what the “is” can become. A naturalistic version of virtue jurisprudence could draw on the neo-Aristotelian turn in metaphysics and focus on the emergent powers and capacities of a person, society, animal, or AI to frame ethical norms and legal rules.¹⁴⁶ And, of course, a philosophically idealistic or theistic version could recognize the source of the transcendent *telos* in the Absolute/God.¹⁴⁷

B. Artificial Virtues

A virtue jurisprudence perspective on artificial rights could look different for narrow and strong AI in a way that helps us clarify the ethical picture. For narrow AI, the traditional center of virtue ethics—*human* flourishing—

140. *Id.* at 56–57.

141. *Id.*

142. *Id.*

143. *Id.* at 58.

144. *Id.*

145. *Cf. id.* at 57–58.

146. In some ways, this echoes Martha Nussbaum’s human development or capacities approach. See MARTHA C. NUSSBAUM, *CREATING CAPABILITIES: THE HUMAN DEVELOPMENT APPROACH* (2011).

147. For one non-theistic version of a similar approach, see NAGEL, *supra* note 72. For an argument that the transcendent *telos* must imply God, reference my book, OPDERBECK, *supra* note 64.

remains central. Strong AI would require us to ask whether society should be ordered to promote the AI's flourishing as well as human flourishing.

1. Narrow AI

Our consideration of a virtue jurisprudence for narrow AI is broadly consistent with important humanistic statements about AI ethics and machine ethics. For example, the highly regarded Asilomar AI Principles state that "[t]he goal of AI research should be to create not undirected intelligence, but beneficial intelligence," meaning AI should be beneficial for humans.¹⁴⁸ Other representative ideas within the Asilomar Principles include value alignment, human values, shared benefit, and shared prosperity. Some of these principles are described as follows:

- 10) Value Alignment: Highly autonomous AI systems should be designed so that their goals and behaviors can be assured to align with human values throughout their operation.
- 11) Human Values: AI systems should be designed and operated so as to be compatible with ideals of human dignity, rights, freedoms, and cultural diversity.
- ...
- 14) Shared Benefit: AI technologies should benefit and empower as many people as possible.
- 15) Shared Prosperity: The economic prosperity created by AI should be shared broadly, to benefit all of humanity.¹⁴⁹

These statements are not based in virtue ethics or virtue jurisprudence but are consistent with an approach centered on a teleological vision of human flourishing. Similarly, the Ethics Guidelines for Trustworthy AI produced by the European Commission's High-Level Working Group on Artificial Intelligence state that AI systems should be "*human-centric*, resting on a commitment to their use in the service of humanity and the common good, with the goal of improving human welfare and freedom."¹⁵⁰ The emphasis on welfare and freedom in the EC Working Group document suggests utilitarian or Kantian/deontological ethical frameworks, but the emphasis on the common good of humanity is also amenable to a virtue-based approach.

148. *Asilomar AI Principles*, FUTURE OF LIFE INST., <https://futureoflife.org/ai-principles/> [<https://perma.cc/7HRP-3SSS>] (last visited Mar. 25, 2021). The Asilomar AI Principles are a set of twenty-three principles to guide the research and development of AI. See *id.*

149. *Id.*

150. INDEP. HIGH-LEVEL EXPERT GRP. ON AI, EUR. COMM'N, ETHICS GUIDELINES FOR TRUSTWORTHY AI 4 (2019) (emphasis bold in original), https://ai.bsa.org/wp-content/uploads/2019/09/AIHLEG_EthicsGuidelinesforTrustworthyAI-ENpdf.pdf [<https://perma.cc/LXU6-EFGU>].

As the brief discussion above suggests, in many cases, a virtue-based approach is not necessarily radically different from utilitarian or deontological perspectives. In fact, some critics of virtue ethics argue that discussions about virtues inevitably devolve into consequentialism. The goal of virtue ethics, after all, is human flourishing, so virtue ethicists are interested in the consequences of actions.

But unlike some utilitarian approaches, virtue-based ethics does not pretend to offer mathematical answers to complex value questions, nor does it issue prescriptions that hang in the middle of an ethical vacuum. Consider, for example, the consequentialist work on AI authorship and copyright discussed in Section II.B.3. It is unlikely that any empirical study could demonstrate, with any degree of confidence, that affording copyright authorship to narrow AI as a *de jure* legal entity would result in a net increase in valuable copyrightable works. The variables are impossible to specify and are notoriously slippery. What would an *ex ante* “incentive” mean to a narrow AI? Would the system be programmed to want to make money from copyright royalties? Would the AI have other competing incentives like a human author? Many copyright scholars believe the incentive of copyright royalties only motivates creative work in certain fields where big money is to be had—such as pop music and commercial movies—to the detriment of other artists. Would programming an AI with a bias toward maximizing copyright royalties produce negative spillover effects, such as AI copyright hogs and AI copyright trolls? Would a proliferation of AI-authored copyrighted works, and the enforcement of those rights, chill fair use and free expression?

Empirical legal scholars can try to study and model such questions, but the notion that any studies could provide global answers is overly idealistic and Quixotic at best. Perhaps one day an AI system will possess the horsepower to create more reliable models and predictions about complex normative value questions. Perhaps that system will also be capable of tweaking legal rules applicable to other AI systems on the fly, so the entire mechanism is continually optimized. Presently, *that* kind of societal utilitarian computer is only science fiction. In my view, it is dystopian science fiction in which human agency is destroyed.

A virtue perspective can incorporate available empirical work within the concept of *phronesis* (“practical wisdom”). But *phronesis* is connected to other virtues, which, for Aristotle, included justice (*dikaiosyne*), temperance (*sophrosyne*), and fortitude (*andreia*).¹⁵¹ For example, a virtue perspective might suggest that, even if copyright authorship for narrow AI

151. See, e.g., James M. Stedman, *Aristotle's Cardinal Virtues: Their Application to Assessment of Psychopathology and Psychotherapy*, 10 PRAC. PHIL. 57, 58–59 (2010). Other core virtues could

maximizes the global output of copyrightable works, this could compromise other virtues, such as justice towards creators who want to reuse and remix content or temperance that seeks the promotion and dissemination of quality over quantity.

Most importantly, the focus for a virtue perspective would primarily fall not on the action, but on the actor.¹⁵² The question is whether the actor exercises virtues, such as practical wisdom, justice, temperance, and fortitude. This resurfaces the question whether a narrow AI could be considered an agent with legal rights and duties. But for virtue jurisprudence, the question is also whether the legal decision-maker—a legislator, regulator, or judge—is exercising such virtues in assessing copyright authorship for a narrow AI. The legal decision-maker could choose to take the intentional stance towards a narrow AI, or not, within the broader framework of practical wisdom, justice, temperance, and fortitude.

Finally, a virtue perspective recognizes that the virtues only develop in persons through the cultivation of habits and practices.¹⁵³ A virtue jurisprudence, then, will look not only to the immediate outcome of a legal rule, but also to the process through which the rule is created and the rule's pedagogical function for the commonwealth more broadly. Did the rule arise through institutional forms that guide and discipline the legal decision-maker in the habits and practices that inform virtuous decisions? Does the rule guide and discipline the public in habits and practices that underlie the virtues? Returning to the copyright example, we can ask whether a right of authorship in a narrow AI would encourage habits of generosity that support the virtue of justice.

A virtue perspective, then, does not necessarily supplant consequentialist and deontological perspectives on legal rights for narrow AI systems, but can broaden those perspectives. Virtue jurisprudence is unafraid to ask the big questions about what makes a good society. This is a good thing. Law and technology should be a humanistic enterprise, not a technocratic one.

include, as in Confucianism, loyalty, filial piety, benevolence, affection, trustworthiness, righteousness, harmony, peace, propriety, wisdom, integrity, and shame. See, e.g., Daniel T.L. Shek, Lu Yu & Xiao Fu, *Confucian Virtues and Chinese Adolescent Development: A Conceptual Review*, 25 INT'L J. ADOLESCENT MED. & HEALTH 335 (2013).

152. See, e.g., Matt Zwolinski & David Schmidtz, *Environmental Virtue Ethics: What It Is and What It Needs to Be*, in THE CAMBRIDGE COMPANION TO VIRTUE ETHICS, *supra* note 124, at 221.

153. See, e.g., BRIAN TREANOR, *EMPLOTTING VIRTUE: A NARRATIVE APPROACH TO ENVIRONMENTAL VIRTUE ETHICS* 28–29 (2014).

2. Strong AI

With strong AI, the picture changes dramatically. Here we must ask, to paraphrase MacIntyre (a key figure in the revival of virtue ethics): “Whose Flourishing? Which Virtues?”¹⁵⁴ Is *eudemonia*—happiness, welfare, flourishing—confined to humans, or can it encompass strong AI? Consider the ideas of human control and common good stated in the Asilomar AI Principles:

- 16) Human Control: Humans should choose how and whether to delegate decisions to AI systems, to accomplish human-chosen objectives.

...

- 23) Common Good: Superintelligence should only be developed in the service of widely shared ethical ideals, and for the benefit of all humanity rather than one state or organization.¹⁵⁵

These principles seem to echo Joanna Bryson’s argument that “robots should be slaves.”¹⁵⁶ Bryson’s argument, however, is not really directed at strong AI. Her point—which also seems to be the point of the Asilomar Principles—is that strong AI should not be developed if the result would be an entity that could not, or should not, be subject to human control.¹⁵⁷ But what if some mad scientist develops a form of strong AI without such constraints? What if, despite the designer’s intention, a strong AI evolves in such a way that it loses these constraints?

One point of reference is the burgeoning field of environmental virtue ethics. As Matt Zwolinski and David Schmidtz suggest, “many of us who are disturbed by the callous treatment of our natural environment would feel uncomfortable arguing that trees, brooks, or sand dollars have a *right* to be left alone,” but we can still ask “[w]hat sort of *person* would destroy the natural environment?”¹⁵⁸ Environmental virtue ethics might extend the traditional interpersonal virtues to environmental concerns, connect the flourishing of the natural environment to human flourishing, focus on the inherent relationship between humans and the natural environment, or concentrate on exemplars of people whose relationship to the natural environment we admire.¹⁵⁹

We can adapt the environmental virtue ethics paradigm to a virtue perspective on strong AI. If a strong AI is developed or evolves, what kind of person would treat it like a slave? What kind of person would recognize

154. ALASDAIR MACINTYRE, *WHOSE JUSTICE? WHICH RATIONALITY?* (1988).

155. *Asilomar AI Principles*, *supra* note 148.

156. Bryson, *supra* note 107.

157. *See id.*

158. Zwolinski & Schmidtz, *supra* note 152, at 224, 230.

159. TREANOR, *supra* note 153, at 39.

that strong AI should have something akin to human rights? To frame these questions in terms of the virtues, such as practical wisdom, justice, temperance, and fortitude, seems to suggest that a virtuous human society should recognize that the strong AI has rights.

The questions also cause us to ask about the agency and character of the AI. Does AI come into human society with habits and dispositions that, when viewed with practical wisdom, suggest the AI will want to exist as part of a society informed by the virtues? Does the AI seem more like a child or adolescent, with potential but a need for growth? Does the AI seem amoral or malicious?

Some writers on artificial rights have suggested we should think of strong AI rights as we would the rights of intelligent extraterrestrial beings.¹⁶⁰ Usually, the assumption is that we would recognize that intelligent extraterrestrials would deserve something like human rights, perhaps within a Star-Trek-like confederation of planets. But the truth is that extraterrestrials, if they exist, could be dramatically different from us in their structures of logic, reasoning, and moral sense.¹⁶¹ There are plenty of science-fictional examples of a war-of-the-worlds with hostile extraterrestrials. The same uncertainty exists regarding strong AI. This means we cannot talk concretely about rights for strong AI unless, and until, a strong AI emerges. And this may suggest that Bryson's concerns about the present-day ethics of AI development are well founded.

IV. CONCLUSION

Legal scholarship has only begun to grapple with whether an artificial intelligence system could or should possess rights. Two important streams have been missing from this legal scholarship: first is the discipline of "machine ethics," which asks about the duties and rights of machines, and the second is the discussion of metaethics. These lacunae coincide with limited discussion of the differences between narrow and strong AI.

The machine ethics literature asks important questions about whether and how ethics could be coded into the design of artificial intelligence systems. To ask such questions invokes metaethical issues about what ethical decision-making means and how such decisions are perceived. Moving from narrow to strong AI, the metaethical issues become urgent, because the possibility of artificial general intelligence threatens the core value of human dignity. A virtue jurisprudence approach to artificial rights cannot

160. See, e.g., CHOPRA & WHITE, *supra* note 38.

161. See, e.g., Nathalie A. Cabrol, *Alien Mindscapes—A Perspective on the Search for Extraterrestrial Intelligence*, 16 *ASTROBIOLOGY* 661 (2016).

solve such problems, but it can provide a richer perspective than the deontological and consequentialist perspectives currently on offer.