

# *Does Artificial Intelligence Require an Artificial Constitution?*

*Conor White\**

## I. INTRODUCTION

Artificial humans are something simultaneously captivating and terrifying, because they are fundamentally familiar yet foreign; a concept that has resulted in a thought experiment considered by individuals for centuries.<sup>1</sup> There exists an issue, however, in this thought experiment—in that it may cease to solely be a thought experiment much longer. With technology developing exponentially and this almost primal fascination with an artificial human being, the chances of there being an artifice that can convincingly portray the human condition continues to increase, and with it a potential issue for the legal community to face: should that creation be granted legal recognition? This is a question that, in some areas of the legal field, has been met with flippant and dismissive responses.<sup>2</sup> However, the question needs to be considered, for the reasons I will explain.

The purpose of this Comment is to focus on personhood and rights for artificial intelligence, and from a foundation of jurisprudence and philosophy, the issue is not as clear cut

---

\*Conor White is a 2021 candidate for Juris Doctorate at SMU-Dedman School of Law. He received a Bachelor of Business Administration from Southern Methodist University in 2018.

1. RENÉ DESCARTES, DISCOURSE ON THE METHOD OF RIGHTLY CONDUCTING ONE'S REASON AND SEEKING TRUTH IN THE SCIENCES, 33 (1637), <https://gallica.bnf.fr/ark:/12148/btv1b86069594/f37.item.zoom>, translated in THE PROJECT OF GUTENBERG EBOOK OF A DISCOURSE ON METHOD 33–35 (Project Gutenberg Org., 2016) <https://www.gutenberg.org/files/59/59-h/59-h.htm>.

2. See, e.g., Lawrence Solum, *Legal Personhood for Artificial Intelligences*, 70 N.C. L. REV. 1231, 1235 (1992); Yoon Chae & Ben Kelly, *Granting AI Legal Personhood Would Be Premature*, LAW360 (May 24, 2019), <https://www.law360.com/articles/1162408/granting-ai-legal-personhood-would-be-premature>.

than first impressions would imply. In this paper, I will have five parts, each addressing their own facet of the argument. Part II will address the history and development of AI, part III will address the argument that AI should be property and why that argument may not suffice, part IV will address the argument that AI is fundamentally not human and why that argument may not suffice, part V will address the argument that AI “lacks” something and why that argument may not suffice, and finally part VI will conclude the paper.

Before continuing, a few concessions and stipulations must be made. Neither this paper, nor I, mean to suggest that the technology addressed in this paper beg an immediate response as they are now. The notion of an artificial person, for the purposes of this argument, requires these things be true: that the AI was developed with the intention of being a human proxy, that the AI can convincingly act in a way that represents a human, and that the AI does function autonomously without oversight or input from the creator. In essence, the AI must convincingly model a fully functional independent adult individual. Since the technology is not yet at this point and much of this paper will revolve around philosophy, this is fundamentally an argument in the abstract—about how things ought to be, or more accurately how they ought to be thought of—and I do not claim it to be anything other than that.

## **II. BACKGROUND AND HISTORY OF ARTIFICIAL INTELLIGENCE**

One of the first to articulate the possibility of machine thinking was, naturally, Descartes, stating that “we can easily understand a machine’s being constituted so that it can utter words, and even emit some responses to actions on it of a corporeal kind,” but asserted that “it never happens that it arranges its speech in various ways, in order to reply appropriately to everything that may

be said in its presence, as even the lowest type of man can do.”<sup>3</sup> Observers have noted that the assertion made by Descartes that a machine could never seamlessly interact with humans is fundamentally at the heart of the AI debate.<sup>4</sup>

In 1950, Alan Turing explored the mathematical possibility of machines using available information and reason to make decisions in his paper *Computing Machinery and Intelligence*.<sup>5</sup> Turing began his discussion succinctly, posing the question “Can machines think?”<sup>6</sup> The term “artificial intelligence” wasn’t officially coined until 1956, at a conference at Dartmouth College at the Dartmouth Summer Research Project on Artificial Intelligence, thus creating the beginnings of the field of AI.<sup>7</sup> While the meeting fell short of organizers’ expectations, the consensus was clear: AI was achievable.<sup>8</sup> What followed this meeting was decades of research and development.<sup>9</sup> One of the organizers of the meeting at Dartmouth, Marvin Minsky, proclaimed in 1970 that a machine with the “general intelligence” of a human being would be available in 3-8 years, but the

---

3. DESCARTES, *supra* note 1 **Error! Bookmark not defined.**, translated in ANIMALS ARE MACHINES 115–118 (London: Cambridge University Press, 1931), <https://webs.wofford.edu/williamsnm/back%20up%20jan%204/hum%20101/animals%20are%20machines%20descartes.pdf>.

4. Solum, *supra* note 2, at 1235.

5. Rockwell Anyoha, *The History of Artificial Intelligence*, SCIENCE IN THE NEWS (Aug. 28, 2017), <http://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/>.

6. Alan Turing, *Computing Machinery and Intelligence*, 49 MIND 433, 433 (1950), <https://www.csee.umbc.edu/courses/471/papers/turing.pdf>.

7. Anyoha, *supra* note 5.

8. *Id.*

9. *Id.*

hurdle of hardware limitations quickly proved to be a difficult obstacle to overcome.<sup>10</sup> Ultimately AI fell out of the public view, but ironically thrived as it turned to private development and the development of hardware allowed for increased computing power.<sup>11</sup> By 1997, IBM's *Deep Blue* had beaten reigning world champion and grand master Gary Kasparov in chess.<sup>12</sup> More recently, IBM developed Watson, a highly powerful AI meant to be more effective in understanding and interacting in natural language, who beat the reigning *Jeopardy!* champion in 2011, further accelerating the development cycle.<sup>13</sup> While Watson was never intended to model the human brain or to interact with language like humans do,<sup>14</sup> other projects like Microsoft's 2016 adventure with the Twitterbot Tay have been developed to do exactly that.<sup>15</sup> While that experience could be categorized as a catastrophic failure for reasons other than computing power or capability,<sup>16</sup> it—along with AI like Siri, Alexa, and Cortana—reflects how far the ability of AI has come in mimicking human interaction.

As noted above, however, we're coming closer and closer to those interactions becoming reality than could have been imagined even twenty-seven years ago.<sup>17</sup> The measure of that

---

10. *Id.*

11. *Id.*

12. *Id.*

13. *A Computer Called Watson*, IBM, <https://www.ibm.com/ibm/history/ibm100/us/en/icons/watson/> (last visited Feb. 8, 2021).

14. *Id.*

15. Asha Barbaschow, *Microsoft and the Learnings From its Failed Tay Artificial Intelligence Bot*, ZDNET (July 24, 2019), <https://www.zdnet.com/article/microsoft-and-the-learnings-from-its-failed-tay-artificial-intelligence-bot/>.

16. *Id.*

17. Barbaschow, *supra* note 15.

capability was first articulated by Alan Turing, who proposed the Turing Test: a series of questioners would ask both a machine and a human a round of questions—on any subject in any way they like—and receive answers from both, without knowing which answers were the machine’s and which answers were the human’s.<sup>18</sup> The heart of the test was not to bring an answer as to whether a machine could *think*, but whether a machine was capable of fooling a series of human questioners that it was in fact a human as frequently as a human answerer—the goal being that the test is so difficult to pass that anything that could pass it must necessarily qualify as intelligent.<sup>19</sup>

That being said, as acknowledged earlier, the technology to reach that point is decisively not at the point of passing that test.<sup>20</sup> So why, then, can the question of personhood for artificial intelligence be asked? Crucially, the rate of technological development increases at an exponential level.<sup>21</sup> Known as Moore’s Law, the generally accepted principle states that the processing speed of computers doubles every eighteen months.<sup>22</sup>

Why this lends credibility to the propriety of this comment’s inquiry is relatively simple: we are getting closer to the realization of human type proxies, and the technology is only going to accelerate its developmental curve.<sup>23</sup> For instance, Boston Dynamics started by developing a

---

18. Solum, *supra* note 2, at 1235–36.

19. *Id.* at 1236.

20. Anyoha, *supra* note 5.

21. *Big Idea: Technology Grows Exponentially*, BIG THINK (Mar. 21, 2011), <https://bigthink.com/think-tank/big-idea-technology-grows-exponentially>.

22. *Id.*

23. *Id.*

quadruped robot that could navigate rough terrain through sensors and a control system in 2004.<sup>24</sup> Sixteen years later, they have developed a bipedal robot capable of jumping, rolling, doing handstands, and generally displaying the ability to move like a human.<sup>25</sup> In fourteen years, IBM went from creating a machine capable of beating a chess grandmaster to creating one capable of besting humans in a general knowledge competition based on conversational question prompts.<sup>26</sup>

At this point, let's return to Microsoft Tay, as was alluded to earlier.<sup>27</sup> Tay was a Twitterbot created by Microsoft intended to engage with a target demographic of eighteen to twenty-four year olds to “engage and entertain people” in a way meant to replicate human conversation.<sup>28</sup> Tay was based on a similar bot developed by Microsoft in China, named “Xiaoice,” who indexed the Bing search engine to datamine it for human conversations in order to identify patterns to replicate in its own speech.<sup>29</sup> Xiaoice reportedly had more than forty million conversations with users in China without incident.<sup>30</sup> Much like its predecessor, Tay used machine learning in order to scan, archive,

---

24. *Legacy Robots*, BOSTON DYNAMICS, <https://www.bostondynamics.com/legacy> (last visited Feb. 8, 2021).

25. *Atlas*, BOSTON DYNAMICS, <https://www.bostondynamics.com/atlas> (last visited Feb. 8, 2021).

26. Anyoha, *supra* note 5; *A Computer Called Watson*, *supra* note 13.

27. Barbaschow, *supra* note 15.

28. *Id.*

29. Caitlin Dewey, *Meet Tay, the Creepy-Realistic Robot Who Talks Just Like a Teen*, WASH. POST (Mar. 23, 2016, 12:32 PM), <https://www.washingtonpost.com/news/the-intersect/wp/2016/03/23/meet-tay-the-creepy-realistic-robot-who-talks-just-like-a-teen/>.

30. Peter Bright, *Tay, the Neo-Nazi Millennial Chatbot, Gets Autopsied*, ARSTECHNICA (Mar. 25, 2016), <https://arstechnica.com/information-technology/2016/03/tay-the-neo-nazi-millennial-chatbot-gets-autopsied/>.

and then use interactions it had with any users that decided to interact with it.<sup>31</sup> Success was almost immediate, amassing a large following and interaction on Twitter in less than twenty-four hours.<sup>32</sup>

For reasons that will be elaborated on later, the Tay experiment was ultimately a failure, but not for reasons that lend themselves to Tay’s inability to communicate.<sup>33</sup> The takeaway here is that the ability for us to develop technology in a way intended to replicate our own *outputs*, regardless of the process or inputs to get there, is advancing rapidly.

The problem of AI does not necessarily lie in the ability to mimic thought, however.<sup>34</sup> The fundamental hang ups regarding the function of AI as human proxies comes from three main objections: that they ought to be property, that AI “lack” something—like intentionality, or that they are fundamentally not human.<sup>35</sup> These three objections constitute the heart of the discussion of this comment, and we’ll begin with the notion that AI ought to be property.

### III. AI IS, OR SHOULD BE, PROPERTY

This argument, for obvious reasons, is initially quite compelling. On its face, the answer is very obvious—AI is a thing created by people for their own use, and as such are simply that—property. But to continue using Tay as a demonstrative example,<sup>36</sup> there are some issues with this

---

31. *Id.*

32. Barbaschow, *supra* note 15.

33. *See id.*

34. *Id.*

35. *See id.*

36. *Id.*

line of thinking as it pertains strictly to artificial intelligence. Almost immediately after launch, Tay’s use of machine learning was exploited, and it was flooded with anti-Semitic, racist, and bigoted hate speech, prompting the machine to use that line of thinking as a basis of output—resulting in a flood of highly controversial tweets from Tay.<sup>37</sup>

The tweets in question related from everything to Holocaust denial, use of slurs to reference President Obama, calling for genocide, and praising Hitler, among others.<sup>38</sup> Many commentators point out, rightfully, that Tay is sending out these tweets without any recognition or understanding of what racism is, and criticized Microsoft for lack of preparation or prevention in anticipation of this happening.<sup>39</sup>

This relates to a concern raised about lack of intentionality—a topic that will be discussed in depth in part V—with regards to AI.<sup>40</sup> Notably, Taylor Swift actually threatened legal action against Microsoft following the incident—after trying to block Microsoft from using the moniker Tay—because of potential confusion and false association between Swift and the chatbot.<sup>41</sup> These

---

37. Rob Price, *Microsoft is Deleting its AI Chatbot’s Incredibly Racist Tweets*, BUSINESS INSIDER (Mar. 24, 2016, 7:31 AM), <http://web.archive.org/web/20190130071430/https://www.businessinsider.com/microsoft-deletes-racist-genocidal-tweets-from-ai-chatbot-tay-2016-3>.

38. *Id.*

39. Price, *supra* note 37; Bright, *supra* note 30.

40. Solum, *supra* at note 2, at 1237.

41. Jennings Brown, *Taylor Swift Threatened Legal Action Against Microsoft Over Racist and Genocidal Chatbot Tay*, GIZMODO, (Sep. 10, 2019, 12:10 PM), <https://gizmodo.com/taylor-swift-threatened-legal-action-against-microsoft-1838011533>.

two things highlight the concern about AI: how do you litigate issues arising from situations like these, issues traditionally only caused and created by people, when they are raised by AI?

Surely, owners of property can be held liable for their property causing harm, whether it is a dog biting someone, injury occurring on premises owned by someone else, or injury caused by property because of negligence of the owner.<sup>42</sup> You could classify these as three different kinds of property as well: animate property (in the case of animals), real property, and personal property.<sup>43</sup> If, at this point, we are inclined to view AI as property—primarily because they are developed by corporate entities or individuals and (at the moment) lack any corporeal vessel convincing enough to pass as persons—there is an issue with those who would be inclined to lean that way.<sup>44</sup>

Up until this point in history, in no way has property been able to exercise any kind of recognized constitutional protection—because there was no property that could actually *exercise* person-like behavior.<sup>45</sup> It is important to note that corporations have received a recognized right to free speech through recent extensions of common law,<sup>46</sup> but it is also important to note that thus far, there is no First Amendment exception to offensive or hateful speech, unless it presents an imminent threat to incite unlawful activity.<sup>47</sup> If corporations have a protection for free speech, and

---

42. *Premises Liability*, JUSTIA, <https://www.justia.com/injury/premises-liability/> (last updated Apr., 2018).

43. *Types of Property for Tax Purposes*, FIND LAW, <https://realestate.findlaw.com/owning-a-home/types-of-property-for-tax-purposes.html> (last updated Sept. 7, 2018).

44. Solum, *supra* note 2, at 1239.

45. *Id.* at 1236.

46. *Citizens United v. Fed. Election Comm'n*, 558 U.S. 310, 342 (2010).

47. *See Matal v. Tam*, 137 S.Ct. 1744, 1763 (2017) (stating that as a matter of law, the Court has consistently held “the public expression of ideas may not be prohibited merely because the ideas

their *property*, designed to act without needing any kind of impetus from the corporation, begins to conduct itself in a way that at least looks like it deserves constitutional protection, how should the law react?

This is precisely the problem: AI like Tay is designed to interact and conduct itself like an employee of the company—that is to say, without requiring action of another employee in order to function—while simultaneously not being employed by that same company.<sup>48</sup> Tay received no compensation, it was not designed to promote Microsoft or act as its agent like Siri does for Apple (otherwise these issues would, presumably, have never occurred), its *sole* purpose was to gather information, identify patterns, and synthesize it in a way that replicated human interaction produced by a machine.<sup>49</sup>

Certain critics look at the Tay incident and put the onus on designers and engineers, and that they should “start thinking about codes of conduct and how accidentally abusive AI can be,”<sup>50</sup> but from a legal context, why? Take the case of *Miller v. Fed. Express Corp.*, a case out of the Indiana Court of Appeals.<sup>51</sup> In this case, a plaintiff sued an employer for the defamatory statements made by an employee while on a company computer.<sup>52</sup> The Court held that, under 47 U.S.C.A. §

---

are offensive to some of their hearers”); *see also* *Texas v. Johnson*, 491 U.S. 397, 414 (1989) (stating that “if there is a bedrock principle underlying the First Amendment, it is that the government may not prohibit the expression of an idea simply because society finds the idea itself offensive or disagreeable.”).

48. Barbaschow, *supra* note 15.

49. *Id.*

50. Bright, *supra* note 30.

51. *Miller v. Fed. Express Corp.*, 6 N.E.3d 1006 (Ind. Ct. App. 2014).

52. *Id.* at 1009.

230, a company providing internet services to employees was a “provider” under the code, the complaint sought to hold the company liable as a publisher, and that the employee poster was “another information content provider,” thus satisfying the three prongs required for protection under the statute.<sup>53</sup> This is hardly the only case holding as such, as the case also cites to authorities from the California Court of Appeals, the Seventh Circuit, and the Illinois Court of Appeals agreeing with the position taken by the court in *Miller*.<sup>54</sup> Given this precedent, and the precedent of free speech under the First Amendment, why is it that engineers or designers need to be responsible for “abuse” caused by AI? Certainly, engineers face liability for failure to design safe buildings, but once again engineers who create a faulty building aren’t creating anything that exercises a constitutionally recognized right.

This precedent creates an issue, because companies like Microsoft who have an AI like Tay may satisfy the three-prong test set out in the code.<sup>55</sup> Tay may very well classify as an “information content provider” as laid out in the Code, and that’s a problem because the Code itself describes an “information content provider” as “any person or entity that is responsible, in whole or in part, for the creation or development of information provided through the Internet or any other interactive computer service.”<sup>56</sup> A company, as a provider, cannot be held liable for information posted by another “information content provider” as part of its provision of internet services, and unless courts intend to distinguish AI from the previous cases and hold companies

---

53. *Id.* at 1018.

54. *Id.* at 1016–18.

55. 47 U.S.C.A. § 230 (2018).

56. § 230(f)(3).

liable for a mostly autonomous piece of software, treating the AI as an “employee” in that case grants it status equal to a “person or entity” under the words of the statute.<sup>57</sup>

This can become even more muddled if we return to the premise of the paper: what about a functional, convincing AI that was created with the intention of being autonomous? At that point, the relation between creation and creator stops looking like an employee and employer or like property and owner. Depending on the role or function intended, it begins to look more like a parental relationship. This statement may seem facetious or off the cuff, but compare the two: children are, by nature, the result of parents creating a human with the intention of that child becoming a fully functioning and independent person.

While the *means* are radically different, the function of the relationship is very similar. If we use this analogy as a means to define the relationship, obviously the view of property fails. Locke rejected the notion of children as property of their parents,<sup>58</sup> and jurisprudence adopts a similar stance. Parents—outside of rare circumstances<sup>59</sup>—are not held liable for the actions of their children, and even then, that duty is removed at the age of majority.<sup>60</sup>

This crossroads presents the conflict underlying the promulgation of AI. If an AI is property, let the company handle it—regardless of the logistics of attaching the AI’s speech to the

---

57. *Id.*

58. Solum, *supra* note 2, at 1277.

59. Janet Portman, *Parents’ Responsibility for Their Children’s Actions*, LAWYERS, <https://www.lawyers.com/legal-info/criminal/juvenile-law/parents-responsibility-for-their-childs-actions.html> (last updated Apr. 9, 2015).

60. *Parental Liability Basics*, FINDLAW, <https://family.findlaw.com/parental-rights-and-liability/parental-liability-basics.html> (last visited Feb. 8, 2021).

appropriately responsible party, whoever that may be—which at this point in time, seems like a simple fix. AI is centered in servers provided by companies, and in most cases, for use by the company or as agents of the company, like Siri, Alexa, or Cortana.<sup>61</sup> But what about AI like Tay, who is simply a company-sponsored semi-autonomous speech bot, or—even more speculatively—when AI can be placed in physical representations of persons? Tay has no function to the company outside of data-gathering to be relayed to the company but acts as a person to those outside the company.<sup>62</sup> Without a way to enforce any duties, as it is now, that creates a tricky problem for the law. You can neither deprive a set of data of liberty, nor does it appear that the company should be held liable for the “speech” of the AI.

This conflict is currently easy to avoid because there is no need to address the possibility, as of right now any and all AI would rightfully classify as property. But when, not if, the ability to produce a functional physical human proxy happens, there *will* be a conflict between the role and relationship of the creator and created. How do you classify the relation between a human and a nonhuman creation?

#### IV. AI IS FUNDAMENTALLY NOT HUMAN

Once again, this is an argument that appears rational, functional, and sensible on its face. I argue, however, that there’s an inherent conflict in the notion embodied by the statement and the treatment of nonhuman entities the law takes. As with the technological side, the jurisprudence

---

61. Randy Bean, *How Big Data is Empowering AI and Machine Learning at Scale*, MIT SLOAN MGMT. REV. (May 8, 2017), <https://sloanreview.mit.edu/article/how-big-data-is-empowering-ai-and-machine-learning-at-scale>.

62. *See* Dewey, *supra* note 29.

recognizing artificial personhood in the US has also seen what could be categorized as a steep development in recent years as well.<sup>63</sup>

In the earliest iteration of the concept of artificial personhood articulated by the Court, Chief Justice John Marshall described a corporation 180 years ago as “an artificial being, invisible, intangible, and existing only in contemplation of the law. Being the mere creature of law, it possesses only those properties which the charter of creation confers upon it, either expressly, or as incidental to its very existence.”<sup>64</sup> Crucially, Marshall’s argument regarding corporate personhood extends beyond reliance on the Framers’ intent in recognizing corporate persons, but rather the applications of the words of the Constitution.<sup>65</sup>

There were limitations, however, on the rights extended to corporate entities.<sup>66</sup> Marshall refused to grant standing to a corporation created by Congress in *Bank of the United States v. Deveaux*, because the corporation was not a state “citizen” within the meaning of the Judiciary Act, stating that “[t]hat invisible, intangible, and artificial being, that mere legal entity, a corporation aggregate, is certainly not a citizen; and, consequently, cannot sue or be sued in the courts of the United States, unless the rights of the members, in this respect, can be exercised in their corporate name.”<sup>67</sup> Marshall was not alone in acknowledging that corporate entities enjoyed constitutional rights, albeit in a limited sense, as Hamilton expressed that certain legal rights and

---

63. Christopher J. Wolfe, “*An Artificial Being*”: *John Marshall and Corporate Personhood*, 40 HARV. J.L. & PUB. POL’Y 201, 201 (2017).

64. *Trs. of Dartmouth Coll. v. Woodward*, 17 U.S. (4 Wheat) 518, 636 (1819).

65. Wolfe, *supra* note 63, at. 210.

66. *Bank of the U.S. v. Deveaux*, 9 U.S. (5 Cranch) 61, 86–87 (1809).

67. *Id.*

duties cannot apply to corporations because of what they are, as they “cannot commit a crime,” or “thought it may be dissolved, it cannot die.”<sup>68</sup> For example, among other restrictions, the Court in *Northwestern Nat. Life Ins. Co. v. Riggs* articulated that “liberty” as protected by the XIV amendment was “the liberty of natural, not artificial persons.”<sup>69</sup>

These limitations, however, began to change and metamorphose very quickly, as the constitutional recognition of those rights extended very quickly.<sup>70</sup> Thirty-five years after the decision in *Deveaux*, the Court ruled that a corporation “created by, and transacting business in a state, is to be deemed an inhabitant of the state, capable of being treated as a citizen, for all purposes of suing and being sued.”<sup>71</sup> Notwithstanding the recent developments in the area of corporate standing,<sup>72</sup> this was among the beginning of the changes to the recognition of constitutional rights for artificial entities. The right of free speech has been recognized as an area of protection for corporate entities.<sup>73</sup> The Court has stated, quite clearly, that “legislature is constitutionally disqualified from dictating the subject about which persons may speak and the speakers who may address a public issue.”<sup>74</sup>

---

68. Wolfe, *supra* note 63, at 206.

69. *Nw. Nat. Life Ins. Co. v. Riggs*, 203 U.S. 243, 255 (1906).

70. *Louisville, Cincinnati & Charleston R.R. v. Letson*, 43 U.S. 497, 497 (1844) (overruled in part by 43 U.S. 497 (superseded by statute by *Hertz Corp. v. Friend*, 130 S.Ct. 1181) (declined to extend by *RJR Nabisco, Inc. v. European Cmty.*, 136 S.Ct. 2090 (2016))).

71. *Id.*

72. *See generally* *Hertz Corp. v. Friend*, 559 U.S. 77 (2010).

73. *First Nat’l Bank of Bos. v. Bellotti*, 435 U.S. 765, 784 (1978).

74. *Id.* at 784–785.

The line of thinking extended by *Bellotti* has created a controversy in artificial personhood jurisprudence.<sup>75</sup> Chief Justice Rehnquist dissented in *Bellotti*—looking to Marshall’s articulation in *Dartmouth College*<sup>76</sup>—stating that “the mere creation of a corporation does not invest it with all the liberties enjoyed by natural persons,” and that the right of political expression cannot be considered “incidental” to a corporation’s very existence.<sup>77</sup> Rehnquist interpreted Marshall’s original belief to be incredibly narrowing, a view shared by many in the Court.<sup>78</sup>

Turn to *Citizens United v. Federal Election Commission*, which reaffirmed *Bellotti*’s position that corporations enjoy a right to free speech, particularly in the realm of political speech.<sup>79</sup> In fact, the Court articulated that First Amendment standards “must give the benefit of any doubt to protecting rather than stifling speech.”<sup>80</sup> However, once again you have a dissenting opinion in Justice Stevens harkening back to Justice Marshall, stating that he found it “implausible that the Framers believed ‘the freedom of speech’ would extend equally to all corporate speakers.”<sup>81</sup>

As a final illustrative case, *Burwell v. Hobby Lobby Stores, Inc.* once again shows an extension of corporate constitutional rights beyond what some justices find conceivable. *Hobby*

---

75. *Bellotti*, 435 U.S. at 822–823 (1978) (Rehnquist, J., dissenting).

76. *See* *Trs. of Dartmouth Coll. v. Woodward*, 17 U.S. (4 Wheat) 518, 636 (1819).

77. *Bellotti*, 435 U.S. at 823–825 (Rehnquist, J., dissenting).

78. *Id.* at 823.

79. *Citizens United v. Fed. Election Comm’n*, 558 U.S. 310, 342 (2010).

80. *Id.* at 327 (quoting *Fed. Election Comm’n v. Wis. Right to Life, Inc.*, 551 U.S. 449, 469 (2007)).

81. *Id.* at 429 (Stevens, J., dissenting).

*Lobby* extended another constitutional protection, in this case freedom of religion, as applied to corporations.<sup>82</sup> The Court held that the “‘exercise of religion’ involves ‘not only belief and profession but the performance of (or abstention from) physical acts’ that are ‘engaged in for religious reasons,’” and that “[b]usiness practices that are compelled or limited by the tenets of a religious doctrine fall comfortably within that definition.”<sup>83</sup> Effectively, companies can freely exercise their own religious beliefs as an exercise of business decisions.<sup>84</sup> This time it was Justice Ginsburg who articulated her own disdain for the extension of the doctrine to corporate rights.<sup>85</sup>

There’s a critical distinction here to be made—companies, while not human, are comprised of humans and often impact the world around them in the way a human might.<sup>86</sup> Companies fundamentally cannot act in a way that a human cannot, because the actors themselves are human.<sup>87</sup> It’s natural to extend rights enjoyed by the individuals that comprise a collective to that collective.<sup>88</sup> The same cannot be said for AI – they are not human and, at the point in which this paper contemplates, will not have any human influence on their actions. This is the basis of the conflict in this argument: how do you approach recognizing something that *functionally* is human but is *fundamentally* not?

---

82. *Burwell v. Hobby Lobby Stores, Inc.*, 573 U.S. 682, 710 (2014).

83. *Id.* (quoting *Emp’t Div., Dept. of Human Res. of Ore. v. Smith*, 494 U.S. 872, 877 (1990)).

84. *Id.* at 713.

85. *Id.* at 751–52 (Ginsburg, J., dissenting).

86. *See id.* at 706.

87. *See id.*

88. *See Hobby Lobby*, 573 U.S. at 706–07.

The first thing to take note of is this: the landscape with regards to both of these areas has changed immensely even in the recent past.<sup>89</sup> In fourteen years, IBM went from creating a machine capable of beating a chess grandmaster to creating one capable of besting humans in a general knowledge competition based on conversational question prompts.<sup>90</sup> As phrased by John Chipman Gray, the “technical legal meaning of a ‘person’ is a subject of legal rights and duties.”<sup>91</sup> One of the arguments against AI’s personhood is that AI itself lacks responsibility.<sup>92</sup> At the same time, at the earliest iteration of corporate personhood, it was also understood that corporations lacked certain capacity for responsibility, in that they could not be held liable for “forfeiture, because it cannot commit a crime,”<sup>93</sup> but recent administrative history knows this all too well to be untrue.<sup>94</sup> The administrative state imposes upon corporations duties while jurisprudence recognizes certain rights that can be exercised by those corporations, and both the set of rights and set of duties have expanded as jurisprudence has become more complex.<sup>95</sup>

So, if we are to entertain the idea of recognizing personhood in AI, how should their rights and duties be defined? Obviously, this depends on how one would choose to classify the

---

89. Anoyha, *supra* note 5.

90. *Id.*; *A Computer Called Watson*, *supra* note 13.

91. Solum, *supra* note 2, at 1238–39.

92. *Id.* at 1244.

93. Wolfe, *supra* note 63, at 206.

94. *See, e.g.*, Ryan Miller, *Johnson & Johnson Ordered to Pay \$8 billion to Man Who Said Drug Caused Him to Grow Breasts*, USA TODAY (Oct. 9, 2019), <https://www.usatoday.com/story/money/2019/10/09/johnson-johnson-8-billion-over-risperdal-gynecomastia-case/3916878002/>.

95. Wolfe, *supra* note 63, at 234–35.

personhood being bestowed. If they are to be recognized as artificial persons, it would make sense to extend to them rights recognized by *some* form of artificial personhood jurisprudence. For that, we have the case law to guide us. But what about duties? Enter Isaac Asimov, who articulated the laws of robotics.<sup>96</sup> Asimov outlines four basic laws of robotics: The Zeroth Law: A robot may not harm humanity, or, by inaction, allow humanity to come to harm.<sup>97</sup> The First Law: A robot may not harm a human being.<sup>98</sup> The Second Law: A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.<sup>99</sup> The Third Law: A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.<sup>100</sup> The lower the number of the law, the higher the priority in the hierarchy of laws.<sup>101</sup> There is already evidence of molecular robots obeying Asimov's Laws of Robotics,<sup>102</sup> so there already exists some precedent that at least in function, some duties can be installed in the processes of AI.

Another area of overlap where corporate personhood at least parallels the discussion on AI is the problem of intentionality. Admittedly, there is a distinction in that corporations are fundamentally comprised of individuals who *do* exhibit intentionality,<sup>103</sup> but in spite of this,

---

96. See ISAAC ASIMOV, FOUNDATION AND EARTH 346 (1986).

97. *Id.* at 347; ISAAC ASIMOV, PRELUDE TO FOUNDATION 397 (1988).

98. ISAAC ASIMOV, LITTLE LOST ROBOT (1947), *reprinted in* I, ROBOT 136, 142 (1950).

99. *Id.*

100. *Id.*

101. *See id.*

102. Gal A. Kaminka et al., *Molecular Robots Obeying Asimov's Three Laws of Robotics*, 23 MIT PRESS 343, 344–45 (2017), [https://www.mitpressjournals.org/doi/full/10.1162/artl\\_a\\_00235](https://www.mitpressjournals.org/doi/full/10.1162/artl_a_00235).

103. Solum, *supra* note 2, at 1239.

jurisprudence still recognizes corporations individually capable of both rights and duties.<sup>104</sup> While many modern judges turn up their noses at the departure from what they believed to be Marshall's original intention in outlining corporate personhood—which some observers have noted may not even be a correct interpretation<sup>105</sup>—acknowledging the classical limitations of corporate personhood means acknowledging a corporation's lack of ability to exhibit intention, as described by both Marshall and Hamilton.<sup>106</sup> That being the case, in spite of that jurisprudence has extended to grant actions by corporations to be wholly and fundamentally attributable to the corporation itself as an entity—regardless of its composition or the individual actors who give it that meaning.<sup>107</sup> That being the case, it becomes harder to make a hard and fast line against intentionality being the *sole* basis for denial of recognition of rights for AI.

Perhaps the actions taken by certain jurisdictions in furthering the notion of artificial personhood can be illustrative. Certain jurisdictions in both the US and internationally have recognized rights for environmental fixtures and landmarks.<sup>108</sup> An organization called the

---

104. *See* *Burwell v. Hobby Lobby Stores, Inc.*, 573 U.S. 682, 710 (2014); Miller, *supra* note 94.

105. Wolfe, *supra* note 63, at 230.

106. *See* *Trs. of Dartmouth Coll. v. Woodward*, 17 U.S. 518, 642–43 (1819); *id.*

107. *See* *Hobby Lobby*, 573 U.S. at 707–08; *Citizens United v. Fed. Election Comm'n*, 558 U.S. 310, 364–65 (2010).

108. *International Movement*, COMMUNITY ENV'T LEGAL DEF. FUND (last updated Nov. 19, 2015), <https://celdf.org/join-the-movement/where-we-work/international-movement/>; *State & National Networks*, COMMUNITY ENV'T LEGAL DEF. FUND (last updated Jan. 26, 2016), <https://celdf.org/join-the-movement/where-we-work/state-national-networks/>.

Community Environmental Legal Defense Fund (CELDF) engages with towns and municipalities to help draft legislation that protects environmental interests from third-parties.<sup>109</sup>

One example of US cities doing this is the town of Shapleigh, Maine.<sup>110</sup> Shapleigh modified its legal code to actually strip some of the rights granted to corporations by the Constitution and granted rights to the nature and natural bodies of water that surrounded the city.<sup>111</sup> The section of the code that defines rights actually grants “unalienable and fundamental rights to exist, flourish, and naturally evolve within the Town of Shapleigh,” and that “ecosystems shall include, but not be limited to, wetlands, streams, rivers, aquifers, and other water systems.”<sup>112</sup>

As an example of the international cases, in New Zealand the national park Te Urewera was named an environmental legal entity, and was made a freehold, inalienable land owned by itself in 2014.<sup>113</sup> New Zealand also named the Whanganui River a legal person in 2017, recognized as “an indivisible and living whole from the mountains to the sea,” incorporating “all of its physical and metaphysical elements.”<sup>114</sup>

---

109. *How We Work*, COMMUNITY ENV’T LEGAL DEF. FUND (last updated Aug. 8, 2019), <https://celdf.org/celdf-services/>.

110. *Water Privatization*, COMMUNITY ENV’T LEGAL DEF. FUND (last updated Dec. 22, 2017), <https://celdf.org/advancing-community-rights/issues/water-privatization/>.

111. SHAPLEIGH, ME, CODE ch., 99, art.11, §§ 99-11, 99-12 (2020), <https://ecode360.com/14302768>.

112. *Id.* § 12.

113. Te Urewera Act, 2014, ss 12-13 (N.Z.).

114. Te Awa Tupua (Whanganui River Claims Settlement) Act 2017, ss 13(b), 14. (N.Z.).

These things being the case—recognition of rights for corporate entities as fundamentally nonhuman “invisible” things and nonliving natural bodies, why *should* the argument that AI is not human be sufficient to claim that personhood is far-fetched? In our hypothetical, AI would be significantly more “person” than either a natural fixture or a corporate entity. While the *premise* of the argument that AI is not human is certainly true, it’s the conclusion that lacks validity.

Solum intimates in his article that the thrust of the nonhuman argument is somewhat fundamentally self-protecting, an almost biological form of xenophobia.<sup>115</sup> He likewise shares my concern that it begins to look more like a *Dred Scott* era way of thinking than anything else<sup>116</sup> — that because something is different from us and as a result scary on a base level, we could and should not afford the protections that we hold dear to us as people to those that are not. However, this also runs into an issue Solum notes: does being a human mean being a person?<sup>117</sup> This is the question that I raise the earlier examples for: while Solum had groundwork for nonhuman personhood to work with in an American context, cases like *Hobby Lobby* had yet to be decided and the notion of personhood for things like a river would have been nothing more than a thought experiment for the sake of the thought experiment—similar to what we’re doing here.<sup>118</sup>

The idea of what a nonhuman “person” would look like and act like is no longer solely the realm of imagination and conjecture, we have programs and AI that can interact and converse

---

115. *See* Solum, *supra* note 2, at 1260–62.

116. *See id.*

117. *Id.* at 1261.

118. *See* Te Awa Tupua Act, ss 13(b), 14; *Burwell v. Hobby Lobby Stores, Inc.*, 573 U.S. 682, 713 (2014).

with us.<sup>119</sup> We now have some kind of foundation for what all this conjecture *could* look like in its fully realized form, and developments in black letter law that push the boundaries of the conception of nonhuman rights.<sup>120</sup>

## V. AI LACK THINGS THAT HUMANS HAVE

This argument is the argument that presents the most turmoil for those who seek concrete answers. It deals largely with the metaphysical—the soul, consciousness, feelings, things that we experience but cannot quantify. As such, it is the most abstract, dealing largely with philosophy of the self, and is also the hardest to overcome by giving examples as well as illustrative examples. What *is* possible to do, and what it is my goal to do, is analyze the argument itself—does an AI *truly* lack what we believe it does, and even if it does, is that enough to satisfy the conclusion that personhood is not warranted?

Let's start with an example of lack: intentionality. Intentionality for the purposes of this objection is the quality of “aboutness.”<sup>121</sup> As John Searle explains, it is the “feature of certain mental states by which they are directed at or about objects and states of affairs in the world.”<sup>122</sup> He explains that “beliefs, desires, and intentions are intentional states; undirected forms of anxiety and depression are not.”<sup>123</sup> One could conceptualize this as the direction of thoughts – how and

---

119. Anyoha, *supra* note 5; *A Computer Called Watson*, *supra* note 13.

120. *See Hobby Lobby*, 573 U.S. at 707–08.

121. Solum, *supra* note 2, at 1267.

122. *Id.* at 1267 n.125.

123. *Id.*

why a thought is directed at something.<sup>124</sup> The basis of this objection is that AI lack the ability to process meaning.<sup>125</sup>

Searle created a thought experiment in response to Turing’s test—known as the Chinese Room—where an individual who does not know Chinese is locked in a room and given batches of Chinese writing from those outside running a version of Turing’s test.<sup>126</sup> That individual is given a rule book outlining a procedure for creating a string of Chinese characters to return to the outside world, convincing the testers that *whatever* is inside the room knows Chinese—in spite of the fact that it does not.<sup>127</sup> The proposition of Searle’s thought experiment is that neither the individual in the room nor the instruction book understands Chinese even though both the individual and the program can simulate such understanding, and by extension a machine running on a program that manipulates symbols in a way that *simulates* human thinking cannot constitute thinking or understanding because the machine lacks intentionality—the ability to process meanings.<sup>128</sup>

While this is most certainly critical in the classical discussion as to whether AI can think, it becomes less substantial in view of whether a machine can be afforded legally cognizable rights—after all, not all rights and duties are subject to intent.<sup>129</sup> Solum recognizes a branch of this argument: if AI cannot live a meaningful life (that is, without the ability to process meaning), does

---

124. *See id.*

125. *Id.* at 1236–37.

126. *Id.* at 1236.

127. Solum, *supra* note 2, **Error! Bookmark not defined.** at 1236.

128. *Id.* at 1236–37.

129. *Id.* at 1239.

that preclude one from attaining personhood?<sup>130</sup> He further goes on to ask how one would test an AI for intentionality, since an intelligent AI would naturally argue they do understand meaning, as it would be critical for certain trials, duties, etc.<sup>131</sup> Because of our experience that only humans, with brains, are capable of understanding, would human factfinders be persuaded that someone lacking a brain could truly process meaning?<sup>132</sup>

Allow me to go one step further in this: why should it matter? It seems to me that the questions and issues some have with the intentionality objection overlooks something: we possess intentionality because experience informs us of the value the information has. What that meaning is will vary from individual to individual depending on their own experiences. This is where the philosophy of this issue can be introduced: the distinction between physicalism and dualism.

Ultimately, the Chinese Room objection and the lack argument presuppose that robots live in a state that embraces strict physicalism: the notion that the real world consists simply of the physical world.<sup>133</sup> This contrasts with the philosophical concept of dualism: that the experience of the world is split into two spheres, the material and the mental.<sup>134</sup> The concern embodied by the lack argument is that human consciousness is fundamentally separated from the processes of the

---

130. *Id.* at 1268.

131. *Id.*

132. *Id.* at 1269.

133. *Physicalism*, DICTIONARY.COM, <https://www.dictionary.com/browse/physicalism?s=ts> (last visited Feb. 8, 2021); *see* Solum, *supra* note 2, at 1239.

134. *Dualism*, DICTIONARY.COM, <https://www.dictionary.com/browse/dualism?s=t> (last visited Feb. 8, 2021).

brain, the *mental states* that we have are distinct from the firing of neurons and processing of information in the brain.<sup>135</sup>

The argument against physicalism can be best conceptualized in a thought experiment promulgated by Frank Jackson, known as the “knowledge argument” or “Mary’s room.”<sup>136</sup> He put forward an idea intended to deny physicalism, it must be false, because “nothing you could tell of a physical sort captures the smell of a rose, for instance.”<sup>137</sup> His argument is this: Mary is a hyper-talented scientist forced to live in a black and white room for the entirety of her life.<sup>138</sup> She knows all there is to know about the physical phenomena of why tomatoes turn ripe, and knows how to use the terms “red” or “blue,” and even knows the specific wavelengths associated with particular shades of a color.<sup>139</sup> She knows how to stimulate the retina and exactly how this produces via the central nervous system the contraction of the vocal cords and expulsion of air to say the words “the sky is blue.”<sup>140</sup> Yet, despite knowing all of this, she herself has never *experienced* the phenomena which she knows how to describe, merely the impetus and naming and quantifiable

---

135. See Solum, *supra* note 2, at 1265–66.

136. See Frank Jackson, *Epiphenomenal Qualia*, 32 *Philosophical Q.* 127 (Apr. (1982)),) *reprinted in* TORIN ALTER & ROBERT J. HOWELL, *CONSCIOUSNESS AND THE MIND-BODY PROBLEM* 23, 25 (1st ed. 2012).

137. *Id.* at 24.

138. *Id.* at 25.

139. *Id.*

140. *Id.*

characteristics of these things.<sup>141</sup> What would happen when she was released from the room and got to experience the world—would she learn anything?<sup>142</sup>

The obvious answer to this question is yes—she would obviously learn what the *experience* of those phenomena are.<sup>143</sup> The issue then, is that her knowledge was incomplete.<sup>144</sup> If we presupposed, however, that she knew *all* physical information there was to know, and physicalism is correct, how is that possible?<sup>145</sup> Thus, Jackson’s conclusion: it is incredibly difficult to deny that one can have all the physical knowledge of something without having all knowledge there is to have.<sup>146</sup> Jackson names that experience, e.g. the seeing of the color red, qualia.<sup>147</sup> Jackson qualifies that this argument is fundamentally different than the question “what is it like to *be*” something—that no amount of information will ever amount to the knowledge of being something you are not.<sup>148</sup>

Compare this to the Chinese Room: in both cases, you have all of the information required to give an answer to something—that is to say in much the same way you know what Chinese characters produce a certain response without knowing what the meanings are, you similarly know

---

141. *See id.*

142. Jackson, *supra* note 136, at 25.

143. *See id.*

144. *Id.*

145. *Id.* at 25–26.

146. *Id.* at 26.

147. *Id.*

148. *See* Jackson, *supra* note 136, at 26–27. (emphasis added).

all of the quantifiable characteristics of a color without actually being able to picture that color.<sup>149</sup> This brings us back to the point I made earlier: humans' intentionality is informed by their *experience* of their information, you like the color red more than blue because you've *seen* it, you understand that.<sup>150</sup>

To extend that to AI, it's almost impossible to say that they would make decisions based on their experiences rather than the parameters set in their programming.<sup>151</sup> Would their experience of a situation inform any future decisions? Intentionality is *qualitative*, and AI is fundamentally *quantitative*, much like Mary in her room.<sup>152</sup> We hesitate to say that AI could possess intentionality because we cannot say what their experience of certain phenomena is like.<sup>153</sup> Because of how AI are created, we confine them to a physicalist approach in life. But recall Jackson's distinction: the rejection of physicalism is distinct from an argument that asks what the experience of *something else* is like.<sup>154</sup> Much like humans' inability to say with certainty that the color "red" one person sees is the same shade of color that another person sees, we *cannot* imagine what the experience of phenomena is to an entity like AI.<sup>155</sup>

---

149. *See id.* at 25–26; Solum, *supra* note 2, at 1237.

150. *See* Jackson, *supra* note 136, at 26–27.

151. *See* Solum, *supra* note 2, at 1265–66.

152. *See* Jackson, *supra* note 136, at 26–27; Solum, *supra* note 2, at 1275–76.

153. *See* Jackson, *supra* note 136, at 27; Solum, *supra* note 2, at 1269.

154. Jackson, *supra* note 136, at 26–27.

155. *See id.* at 24; Solum, *supra* note 2, at 1236–37.

The basis of the intentionality objection issue rests on a few presumptions: that intentionality is more than the physical, that the AI experience is fundamentally physicalist, and without qualia one cannot have intention in their decisions. Put more succinctly, it's easy to say that AI lacks intentionality when you draw these conclusions because the answer presupposes itself. But, as with a lot of this discussion, we don't *know* that the second two presumptions are true.

Even if AI is to lack human qualities that make the human experience fundamentally unique, is that enough to say that personhood is precluded? If we return to the intentionality issue, if an AI was on trial for something that required intention, would an argument of "because my programming is not designed to make me act that way" really not suffice? The framework of an AI's programming doesn't necessarily come out in a binary decision-making tree. At this level of complexity, certain considerations and factors would all have to be weighed and considered, and the *outcome* is the intention less than the process. If one could argue that the parameters of the AI don't tend towards the undesired outcome, that should be enough to satisfy that the act was not intentional. It's not a substitution for *mens rea*, but it could function the same way.

The question of the intentionality issue betrays the foundation for asking it. If we are to believe that an AI would be incapable of committing fraud because they would lack the intention to falsely induce someone into a transaction, they would never need to go on trial.<sup>156</sup> A person incapable of committing a crime doesn't raise an issue of an inability to enforce rights and duties, it simply makes a person incapable of breaching that duty.<sup>157</sup> Likewise, if there is a legitimate

---

156. Solum, *supra* note 2, at 1248.

157. *Id.* at 1244–45.

concern that AI *could* commit that crime, that would also necessarily mean that AI is capable of intentionality.<sup>158</sup>

Intentionality, and anything else that could be thrown under a lack argument, are terms and categories that are human creations.<sup>159</sup> As such, the understanding of the term can change. If we are concerned that an AI could make an illegal decision that requires intentionality, then they are attributed with the characteristic of intentionality.<sup>160</sup> If we are unconcerned about that possibility, then they lack intentionality, but simultaneously absolve themselves of needing that duty enforced.<sup>161</sup> There's an inherent hypocrisy in concerning ourselves with the possibility that something could be created without the ability to break certain laws and also the need to enforce those laws against them.

Let's turn now to the argument that AI lack consciousness. This discussion will bring us even further down this philosophical rabbit hole. Some may not be satisfied with a simple offer of the *explanation* of dualism as the human experience as a merit for recognizing these issues AI present.<sup>162</sup> Even if we settle that intentionality would be immaterial to the recognition of rights, AI isn't conscious as we know it.<sup>163</sup> So – the philosophical rebuttal: while the natural tendency may

---

158. *Id.* at 1267.

159. *Id.* at 1262.

160. *Id.* at 1267.

161. *Id.*

162. Solum, *supra* note 2, at 1263.

163. *Id.* at 1283.

be to tend towards dualism as a way to explain the experiential form of consciousness, there's an inherent problem with it.<sup>164</sup> This brings us to the mind-body problem.

As it stands, dualists use an argument that can be referred to as the “Disembodiment Argument,” a derivative of Descartes’ initial arguments, to explain the relation between body and mind.<sup>165</sup> It focuses on pain, a purely experiential phenomenon, and posits, in effect, the following argument: (1) I can conceive of pain at this moment right now absent any physical cause for it, (2) if I can conceive of a scenario, that scenario is possible, (3) so it is *possible* that this pain exists in a purely disembodied pain, and if it is, it cannot exist in a physical thing, and (4) that being the case, this very pain is not identical to any physical state, therefore (5) the argument that every mental state is linked to a physical state is false.<sup>166</sup>

There is a conflict between our understanding of science and the recognition of the dualist version of consciousness—that is, that consciousness is something separated entirely from the physical—called epiphenomenalism.<sup>167</sup> Boiled down, the rejection is that if there is truly a part of the mind that is completely distinct from the physical form, then there is no possible way that the mind can create an impetus for moving the body.<sup>168</sup> Gertler responds that there is a conflict in our

---

164. *Id.* at 1263.

165. BRIE GERTLER, IN DEFENSE OF MIND-BODY DUALISM, *reprinted in* TORIN ALTER & ROBERT J. HOWELL, CONSCIOUSNESS AND THE MIND-BODY PROBLEM 37 (2012).

166. *Id.*

167. *Id.* at 42.

168. *Id.*

“completeness” of understanding causation, that just because something *regularly precedes* an outcome does not necessarily mean that the former caused the latter.<sup>169</sup>

The problem of mental causation can be discerned from an apparent inconsistency among four propositions.<sup>170</sup> The four propositions are as follows:

- 1) Mental Distinctness: the mental is not identical with the physical.
- 2) Physical Adequacy: physical events have sufficient physical causes if they are caused at all.
- 3) Mental Causation: Some physical events are caused by mental events.
- 4) Non-overdetermination: Not every case of mental causation is a case of over-determination.<sup>171</sup>

Mental causation presents a serious issue with those who *do* adopt the dualist approach to these arguments.<sup>172</sup>

David Chalmers articulated it this way: our claims and judgments about consciousness can be explained in terms independent of consciousness.<sup>173</sup> Ergo, consciousness is *explanatorily*

---

169. *Id.* at 43.

170. TORIN ALTER AND ROBERT J. HOWELL, CONSCIOUSNESS AND THE MIND BODY PROBLEM 350 (2012).

171. *Id.*

172. *Id.*

173. DAVID J. CHALMERS, THE PARADOX OF PHENOMENAL JUDGMENT, *reprinted in* TORIN ALTER AND ROBERT J. HOWELL, CONSCIOUSNESS AND THE MIND BODY PROBLEM 369 (2012).

*irrelevant* to the claims and judgments about consciousness.<sup>174</sup> This begins to look once again like Mary's room: does the ability to describe and explain certain qualia independent from the experience of the qualia itself?<sup>175</sup> Chalmers calls this quirk with consciousness the "paradox of phenomenal judgment."<sup>176</sup> He notes two observers that have commented on the paradox, and notes that one suggests we should be reconciled to this paradox.<sup>177</sup> In the commentators words,

[n]o analysis of the purely physical processes in a brain (or a computer) seems capable of capturing the particular quality of the subjective experience corresponding to those processes. Yet, some such analysis should surely be able to give a causal account of how an individual comes to type a sentence such as the preceding.<sup>178</sup>

So Chalmers continues that while the question of whether consciousness is *causally* relevant is distinct from whether it is *explanatorily* distinct.<sup>179</sup> There exists a way to give a physical explanation of behavior without *any* need to invoke thoughts of consciousness or the metaphysical.<sup>180</sup> As such, explanations of this physical kind will apply equally to "a zombie as to an honest-to-goodness conscious experience."<sup>181</sup>

---

174. *Id.* at 369–70.

175. *See* Jackson, *supra* note 136, at 24.

176. CHALMERS, *supra* note 173, at 370.

177. *Id.*

178. *Id.*

179. *Id.*

180. *Id.*

181. *Id.*

In proposing this, Chalmers explains that retelling a story about the interactions of memories effectively retells the physical story at a higher level of abstraction.<sup>182</sup> Put more simply, the retelling of the memories will omit more details—clutter—by giving an upper-level explanation that we can readily relate to.<sup>183</sup> In retelling the story, saying “the ball moved past me like a blur” might be more satisfying in retelling, but it could just the same be rendered an irrelevant part of the explanation by instead offering the mechanics of how the ball was moving that way.<sup>184</sup> In explaining the story from the low-level physical, however, there might necessarily be an implication of the higher-level phenomena.<sup>185</sup> In these situations, Chalmers explains, that is a result from their logically supervenient status—that is to say the higher-level abstraction is a logical addition to the physical explanation—so there is no problem of explanatory irrelevance, the problem Chalmers identified with consciousness.<sup>186</sup>

For consciousness, there is no such logical supervenience on the physical—otherwise this discussion would not be necessary—so there is a very serious problem with the consciousness discussion.<sup>187</sup> The reason is this: an explanation of behavior can be given entirely in terms that don’t even imply the existence of consciousness.<sup>188</sup> This is not to say that you can *never* invoke

---

182. CHALMERS, *supra* note 173, at 371.

183. *Id.*

184. *See id.*

185. *Id.*

186. *Id.*

187. *Id.*

188. CHALMERS, *supra* note 173, at 371.

consciousness in an explanatory sense, just that it is not *necessary*.<sup>189</sup> That necessity is the crux of the problem.

Chalmers puts forward a hypothetical that is particularly apropos in our context: returning to his zombie mentioned earlier, suppose that zombie talks about his conscious experience all the time, the joys he gets from seeing purples and greens, the qualia of his life.<sup>190</sup> We know from our premise that he necessarily lacks consciousness that he doesn't have the experiences in the way we anticipate, yet there must be *some* explanation for the claims that he makes.<sup>191</sup> Those explanations for his claims must then *necessarily* exist in terms of physical processes and laws, because those are the *only* processes and laws that exist in his world.<sup>192</sup>

Suppose, further, that the same zombie makes the same arguments mentioned earlier—lamenting the inability of some other zombie in some other universe who lacks consciousness, while maintaining that he himself has it, something we know to be abjectly false.<sup>193</sup> So how do you reconcile knowledge of consciousness with the fact that consciousness is explanatorily irrelevant to phenomenal judgments? The paradox, finally, can be ultimately summed up with another argument offered in premise form:

- 1) The physical domain is causally closed;

---

189. *Id.*

190. *Id.*

191. *Id.* at 372.

192. *Id.*

193. *Id.*

- 2) Judgments about consciousness are logically supervenient on the physical;
- 3) Consciousness is not logically supervenient on the physical; and
- 4) We know we are conscious.<sup>194</sup>

Explained in English: from premises one and two, we know that judgments about consciousness can be reductively explained, and in conjunction with premise three, that consciousness is explanatorily irrelevant to our judgments.<sup>195</sup> These taken together, ultimately, are in conflict with premise four.<sup>196</sup> That, ultimately, is the heart of the problem.<sup>197</sup>

So, following the very long-winded discussion of consciousness, how does it all relate to the topic *we're* interested in? Reconsider the hypothetical proposed by Chalmers: of the zombie who lacks consciousness that is fixated on his own conscious experience of life, explaining things and their impact on him<sup>198</sup>—this echoes almost exactly the hypothetical world in which we are envisioning. If one of the main rejections of AI personhood is that they lack a consciousness, how then do we reconcile that as a basis for rejecting legal recognition of something that can interact, explain, and *convey* an understanding of the world in a way indistinguishable from a human? That is to say, if without prior knowledge that the individual conveying these feelings was a construct made of hardware and electrical circuits you would be unable to say that it was not a person, why

---

194. CHALMERS, *supra* note 173, at 374.

195. *Id.*

196. *Id.*

197. *Id.*

198. *Id.*

does that knowledge fundamentally change the approach we would have for such a construct? Fear is the most forthright explanation, but fear alone cannot justify such an explanation.

This is compounded by the inherent issues highlighted in the consciousness arguments: we don't *truly* know what consciousness *is*.<sup>199</sup> Spectators acknowledge that we, as individuals, will never be able to tell the difference between person one's red and person two's red, if such a difference exists.<sup>200</sup> This harkens back to one of Jackson's fundamental issues: we cannot in any way be able to explain the experience of phenomena in a way we do not personally experience.<sup>201</sup> So while the knowledge that as a society we all have certain thoughts or experiences in a particular way, that is there are aspects of our lives that escape immediate physical reduction, like pain, and that all members of that society are human lead us to the conclusion that all of us share a common, unique trait of consciousness,<sup>202</sup> is it really possible to say that AI would not have, in some form, consciousness?

This possibility becomes all the more real and terrifying if we do accept the materialist view of the world. Presume that each and every thought, experience, and phenomenal state *are* couched entirely in the physical—all of it is a result of the processes of our brain and no other extraordinary impulses.<sup>203</sup> After all, our understanding of physics is complete, at least in the

---

199. Solum, *supra* note 2, at 1264.

200. TERENCE HORGAN, JACKSON ON PHYSICAL INFORMATION AND QUALIA, *reprinted in* TORIN ALTER & ROBERT J. HOWELL, CONSCIOUSNESS AND THE MIND-BODY PROBLEM 77 (2012).

201. Jackson, *supra* note 136, at 26–27.

202. *Id.* at 26.

203. GERTLER, *supra* note 165, at 42.

concept of its ability to define the physical processes of the world.<sup>204</sup> If we are to accept this as true, why then must it be the case that another entity who can articulate and explain the world like we do lack the ability to *experience* the way we do?

The argument that AI lacks consciousness as a means to deny personhood *rests* on the presumption that human consciousness is fundamentally non-physical.<sup>205</sup> If it *is* physical, as our scientific understanding of things at the moment suggest,<sup>206</sup> then it would be untenable to suggest that as a matter of certainty AI could not possibly have a consciousness. This then begins to look like a problem of rhetoric: we define consciousness as something metaphysical out of an inability to describe or define the parts of our experience that defy physical sources, but that connection is explanatorily irrelevant, as noted by Chalmers.<sup>207</sup> We also acknowledge that while the *causation* is subject to scrutiny as to what is metaphysically possible,<sup>208</sup> our current scientific understanding denies the conclusion that something metaphysical can exert any kind of influence on the physical.<sup>209</sup>

As a conclusion of all of these factors, the argument that AI lacks consciousness as a basis for denial of legal recognition must fail. Either we cannot say with any certainty that the statement is true, or if it is, it is immaterial for the purposes of the way the law functions. Consciousness is

---

204. *Id.*

205. *Id.*

206. *See* ALTER, *supra* note 170, at 350.

207. CHALMERS, *supra* note 173, at 370.

208. *Id.*

209. ALTER, *supra* note 170, at 349.

fundamentally an individual's experience of the world around them, and the law is fundamentally a guideline for how individuals interact with one another in the world—the law does not require consciousness as a prerequisite, just understanding of the rules.<sup>210</sup> As a matter of point, until this era there has existed no thing that could *understand* the law without—potentially—lacking consciousness.

Solum notes other areas of lack that may constitute rejection of consideration for the premise I suggest, among them are things like lack of feelings, souls, interests, and free wills.<sup>211</sup> The first two are rather easy to dismiss if you are to take my arguments earlier. Lack of feeling does not, in my mind, constitute a tenable argument as a basis for rejection of legal personhood. While correctly acknowledging this is a fundamental difference between a person and an AI, we return to my initial skepticism about how we define the phenomenal: how are we to *know* that they can't feel, else why does it matter if they can *convey* that they do? Additionally, there are humans that lack certain emotion, and that in and of itself does not make them legally any less people than others.<sup>212</sup> Souls are like consciousness, incredibly hard to describe, define, or even verify, and as such my arguments there apply in much the same way here.<sup>213</sup>

---

210. Solum, *supra* note 2, at 1222–23.

211. *Id.* at 1262, 1269, 1271–72.

212. See Rhonda Freeman, *Do Psychopaths Have Emotions?*, NEUROINSTINCTS, <https://neuroinstincts.com/do-psychopaths-have-emotions/> (last visited Feb. 8, 2021).

213. Solum, *supra* note 2, at 1262–63.

Interests—that is to say, the conception of a good life—seems an interesting area.<sup>214</sup> But ultimately, like me, Solum comes to the conclusion that given the variety of society’s views of “good,” that “particular conceptions of the good do not provide an appropriate or even feasible standard for the resolution of the legal question whether AIs are entitled to the rights of constitutional personhood.”<sup>215</sup> Finally, as for the free will argument, it is automatically discounted by one of the very premises of my initial argument—to even get to the point where asking this question would be feasible, we must presuppose that the AI in question can and do function autonomously as their own independent entities. As a result, at the point in time at which I aim to suggest this question is valid this objection will already be moot.

## VI. CONCLUSION AND APPLICATION

We have covered an extensive amount of information in this paper—some straightforward and some very abstract. But taken in its totality, I maintain that the question posed regarding personhood for AI is one that *will* need to be considered. Let’s reestablish the basic premises of the kind of AI I imagine this discussion taking place around: that the AI was developed with the intention of being a human proxy, that the AI can convincingly act in a way that represents a human, and that the AI does function autonomously without oversight or input from the creator. We already have hardware capable of moving in a way very similar to a human.<sup>216</sup> While there

---

214. *Id.* at 1271–72.

215. *Id.* at 1272.

216. *See Atlas, supra* note 25.

will still need to be strides to make it *appear* more like a human, the developments are rapid. In the span of a few decades, we have seen massive jumps in software and functionality of AI.<sup>217</sup>

Tay presented a situation in which we saw that AI *can* fundamentally interact with humans in a way that other humans might.<sup>218</sup> What Tay also taught us, however, is that there is a long way to go with that technology—that at this early incubation stage is subject to the ability to be hijacked and flooded with inputs that create outputs nobody wants.<sup>219</sup> But does this not reflect—to some extent—human behavior? Just like anyone else, without proper guidance those same kinds of ideals can be transferred to humans in much the same way as the internet did with Tay.

Given the fact that we are seeing prototypes that have already far exceeded what most people could have anticipated even two decades ago, and coupled with the fact that this will only continue to develop faster and faster,<sup>220</sup> we need to be conscious of this. More importantly, we need to be critical of the traditional arguments against the notion of personhood for AI. The fact that AI is not human does not seem to be tenable among the current state of law. With companies being able to exercise free speech and religion as an *entity* rather than as people who make up that entity, we have already made steps towards removing that requirement for the purpose of law.<sup>221</sup> This is further buttressed by the acknowledgement that there have been steps made in multiple

---

217. *See A Computer Called Watson, supra* note 13.

218. *See Dewey, supra* note 29.

219. *See Price, supra* note 37.

220. *Big Idea: Technology Grows Exponentially, supra* note 21.

221. *See Burwell v. Hobby Lobby Stores, Inc.*, 573 U.S. 682, 710 (2014).

places to grant legal personhood to natural fixtures out of a recognized need to protect them<sup>222</sup>—an idea and policy that I believe could rightly be extended to the type of AI I premise these arguments on.

As far as the notion that AI should be considered property, there are shortcomings inherent in the concept from a legal standpoint as noted earlier.<sup>223</sup> Tay—even in her rudimentary form—may very well satisfy the exception for liability for Google: then what?<sup>224</sup> If you can't bring recourse against the owner of the property for things like that, doesn't the basic value of classifying the AI as property—to recoup some loss as a result of the AI's conduct from the owner—already fall short? Extend that to the scenario I've envisioned: if the autonomous AI have been created with the goal of being just that—autonomous—does the preconception that the AI is property not fall short to a certain extent already? Presumably at that point, the autonomous AI would be able to live in a way such that it itself has property or assets (assuming of course it was designed to behave like a human, one of the basic premises here), and as such bringing suit against it would not be a zero sum game.

It is the final point of discussion that I believe is the hardest for people to get over, but it is also the point that brings the need for this discussion. Up until this point in history, persons were persons because they were human. The question we need to ask ourselves as we approach the hypothetical scenario is if that is sufficient. All of the rejections and arguments presented under Part V do no more than highlight what would make a robot different from a human—and even

---

222. *See* SHAPLEIGH, ME., CODE §§ 99-11, 99-12 (1997); Te Urewera Act 2014, ss 12-13; Te Awa Tupua (Whanganui River Claims Settlement) Act 2017, ss 13(b), 14.

223. *See* 47 U.S.C.A. § 230(c)(1) (West 2018).

224. *Id.* § 230(f)(3).

then, there are certain shortcomings that I highlighted.<sup>225</sup> It will take a fundamental reevaluation of what constitutes a person in the eyes of the law: is it something that behaves as a member of society, interacting and behaving like the other members that make up the persons in it, or is a person a *human*? While there are outlets to say that there are certain rights available only to *natural* persons,<sup>226</sup> what do we do with unnatural persons who act and behave like natural ones? This is the conflict we have to address with this question: with as little as we *really* know about the human experience, from how we define phenomena,<sup>227</sup> to what consciousness really is,<sup>228</sup> can we morally say that these unnatural persons, despite largely behaving the same as humans, don't require the same kind of protection as us? Out of hesitance to put form over function, and a recognition that this same kind of thinking once applied to slaves,<sup>229</sup> I am not confident that we can. Thus, we're forced with a dilemma: either we prevent development from reaching the point at which these premises come true, or we address the elephant in the room when they do.

---

225. *See supra* Part V.

226. *See* Trs. of Dartmouth Coll. v. Woodward, 17 U.S. (4 Wheat.) 636 (1819).

227. *See* Jackson, *supra* note 136, at 24.

228. *See* CHALMERS, *supra* note 173, at 369.

229. *See* Solum, *supra* note 2, at 1261.