

2018

The Bot Legal Code: Developing a Legally Compliant Artificial Intelligence

Edmund Mokhtarian

Follow this and additional works at: <https://scholarship.law.vanderbilt.edu/jetlaw>



Part of the [Computer Law Commons](#), [Internet Law Commons](#), and the [Science and Technology Law Commons](#)

Recommended Citation

Edmund Mokhtarian, *The Bot Legal Code: Developing a Legally Compliant Artificial Intelligence*, 21 *Vanderbilt Journal of Entertainment and Technology Law* 145 (2020)
Available at: <https://scholarship.law.vanderbilt.edu/jetlaw/vol21/iss1/3>

This Article is brought to you for free and open access by Scholarship@Vanderbilt Law. It has been accepted for inclusion in *Vanderbilt Journal of Entertainment & Technology Law* by an authorized editor of Scholarship@Vanderbilt Law. For more information, please contact mark.j.williams@vanderbilt.edu.

The Bot Legal Code: Developing a Legally Compliant Artificial Intelligence

*Edmund Mokhtarian**

ABSTRACT

The advent of sophisticated artificial intelligence (AI) agents, or bots, raises the question: How do we ensure that these bots act appropriately? Within a decade, AI will be ubiquitous, with billions of active bots influencing nearly every industry and daily activity. Given the extensiveness of AI activity, it will be nearly impossible to explicitly program bots with detailed instructions on permitted and prohibited actions, particularly as they face unpredictable, novel situations. Rather, if risks to humans are to be mitigated, bots must have some overriding moral or legal compass—a set of “AI Laws”—to allow them to adapt to whatever scenarios they face.

This Article demonstrates how to develop a “Bot Legal Code,” a system of AI Laws that can ensure AI compliance with legal (but not moral) requirements. Most proposals for AI Laws—such as Isaac Asimov’s Laws of Robotics—have emphasized that bots must be moral or “good,” but moral precepts operate at a level of abstraction that computers simply cannot grasp. In contrast, a bot can understand legal requirements specifically because the law eschews abstraction in a variety of ways, including through a rich history of case law and rules of conflict resolution. In making this argument, this Article draws a parallel between AI architecture (on the one hand) and legal rules and standards (on the other), demonstrating that AI architecture is already optimized for understanding rules through explicit coding and standards through data processing. This Article then describes the ideal qualities for the Bot Legal Code and addresses how government and peer production communities can develop open-source software to implement the Bot Legal Code.

* Associate, DLA Piper LLP (US); J.D., Harvard Law School, 2013. The views contained in this Article represent the personal views of the Author and not the views of DLA Piper LLP (US), any of its clients, or any of its affiliated international firms or their clients.

TABLE OF CONTENTS

I.	INTRODUCTION	147
II.	AI CONTROL AND AI LAWS	152
	A. <i>The AI Control Problem</i>	153
	B. <i>AI Laws and Ethics</i>	157
	1. Isaac Asimov's Laws of Robotics.....	158
	2. The European Parliament's General Principles of Robotics	161
	3. Oren Etzioni's AI Rules.....	164
	4. Google's Technical Guidelines for Intelligent Agents	167
	C. <i>Conclusion: Revisiting the Proposals of AI Laws</i>	170
III.	MORALITY AS A POTENTIAL DOMAIN OF AI LAWS	171
	A. <i>Benefits of Moral AI Laws</i>	172
	B. <i>Downsides of Moral AI Laws</i>	173
	C. <i>Relative Morality: A Different Form of Moral AI Laws</i>	174
	D. <i>Conclusion</i>	177
IV.	A COMPUTING MODEL OF LEGAL ABSTRACTION: MAKING LAWS MACHINE INTERPRETABLE	178
	A. <i>How Bots Think: A Computer's Instruction and Data Components</i>	179
	1. Explicit Commands Through the Instruction Component	179
	2. Learning Through the Data Component	181
	B. <i>How Laws Operate</i>	184
	C. <i>Encoding Laws into a Machine</i>	188
	D. <i>Conclusion: The Path to the Bot Legal Code</i>	191
V.	THE BOT LEGAL CODE.....	192
	A. <i>Ideal Qualities in the Bot Legal Code</i>	192
	1. Modularity.....	193
	2. Depth.....	194
	3. Consistency	195
	4. Conflict Resolution.....	196
	B. <i>The Open-Source Solution</i>	198
	C. <i>Government and Peer Production</i>	202
	D. <i>Open-Source Software: Implications for the Bot Legal Code</i>	205
VI.	CONCLUSION	206

I. INTRODUCTION

The advent of sophisticated artificial intelligence (AI)¹ programs raises the question: how do we ensure that these programs act appropriately? In the span of one decade, we have developed bots² that enhance driving,³ diagnose cancer,⁴ beat world champions at poker⁵ and Go,⁶ and research answers to basic questions.⁷ In coming decades, commentators predict that such agents will be able to perform medical

1. For clarity, the term “artificial intelligence,” or “AI,” will be used either to denote (1) the field of artificial intelligence (i.e., the “branch of computer science dealing with the simulation of intelligent behavior in computers”) or (2) a sophisticated machine’s ability to approximate human-like intelligence. See *Artificial Intelligence*, MERRIAM-WEBSTER, <https://www.merriam-webster.com/dictionary/artificial%20intelligence> [<https://perma.cc/9ZTB-B5TV>] (last visited Sept. 3, 2018). However, the term “AI” is also used colloquially to refer to AI-powered programs. Devin Coldewey, *‘AI-powered’ is Tech’s Meaningless Equivalent of ‘All Natural’*, TECHCRUNCH (Jan. 10, 2017), <https://techcrunch.com/2017/01/10/ai-powered-is-techs-meaningless-equivalent-of-all-natural/> [<https://perma.cc/4QDH-2UKQ>].

2. A “bot” is any artificially intelligent program, agent, or application—whether a chatbot (text-based digital agent), voicebot (voice-based digital agent), robot, or otherwise. See *Bot*, MERRIAM-WEBSTER, <https://www.merriam-webster.com/dictionary/bot> [<https://perma.cc/7T9Q-25YM>] (last visited Sept. 3, 2018).

3. See, e.g., *10 Million Self-Driving Cars Will Be on the Road by 2020*, BUS. INSIDER (June 15, 2016, 7:25 AM), <http://www.businessinsider.com/report-10-million-self-driving-cars-will-be-on-the-road-by-2020-2015-5-6> [<https://perma.cc/FXL9-ET6Q>].

4. See, e.g., Jon Fingas, *IBM’s Watson AI Saved a Woman from Leukemia*, ENGADGET (Aug. 7, 2016), <https://www.engadget.com/2016/08/07/ibms-watson-ai-saved-a-woman-from-leukemia/> [<https://perma.cc/E7MC-BHHY>]; Sy Mukherjee, *This New AI Can Detect a Deadly Cancer Early With 86% Accuracy*, FORTUNE (Oct. 30, 2017), <http://fortune.com/2017/10/30/ai-early-cancer-detection/> [<https://perma.cc/8Z5L-D8LR>].

5. See, e.g., Cade Metz, *Inside Libratus, the Poker AI that Out-Bluffed the Best Humans*, WIRED (Feb. 1, 2017, 7:00 AM), <https://www.wired.com/2017/02/libratus/> [<https://perma.cc/E7R5-JLYZ>].

6. See, e.g., Christopher Moyer, *How Google’s AlphaGo Beat a Go World Champion*, ATLANTIC (Mar. 28, 2016), <https://www.theatlantic.com/technology/archive/2016/03/the-invisible-opponent/475611/> [<https://perma.cc/HB96-7NDF>]. For the article describing how AlphaGo learned to play Go, see generally David Silver et al., *Mastering the Game of Go with Deep Neural Networks and Tree Search*, 529 NATURE: INT’L J. SCI. 484 (Jan. 1, 2016).

7. See, e.g., Cade Metz, *Google’s Hand-Fed AI Now Gives Answers, Not Just Search Results*, WIRED (Nov. 29, 2016, 7:00 AM), <https://www.wired.com/2016/11/googles-search-engine-can-now-answer-questions-human-help/> [<https://perma.cc/V7ZY-BXMD>].

treatment and surgery,⁸ cook,⁹ clean,¹⁰ fly,¹¹ write,¹² and provide cybersecurity,¹³ to name only a few areas. Bots will be everywhere and embedded in a wide range of devices including phones, computers, home appliances, robots, and augmented reality glasses.¹⁴

Moreover, these bots may be largely uncontrollable.¹⁵ There are already indications of potential problems managing sophisticated

8. See, e.g., Sveta McShane, *The Future of Surgery Is Robotic, Data-Driven, and Artificially Intelligent*, SINGULARITY HUB (Oct. 11, 2016), <https://singularityhub.com/2016/10/11/the-future-of-surgery-is-robotic-data-driven-and-artificially-intelligent/#sm.000t2wraf1306fo6rzd20ac545jib> [<https://perma.cc/B2ZK-Y5XA>].

9. For early signs of such technology, see IBM's Chef Watson application, which builds custom recipes by analyzing ingredient synergy. See, e.g., Jonathan Amos, *'Robot Chef' Aimed at Home Kitchen*, BBC (Apr. 14, 2015), <http://www.bbc.com/news/science-environment-32282131> [<https://perma.cc/XXR8-KSP3>] (showing how a robot chef learns by watching humans cook); Rochelle Billow, *How IBM's Chef Watson Actually Works*, BON APPÉTIT (June 30, 2014), <https://www.bonappetit.com/entertaining-style/trends-news/article/how-ibm-chef-watson-works> [<https://perma.cc/VM4F-97D6>].

10. See, e.g., Madhav Srinath, *Artificially Intelligent Homes: The Robot Vacuum Cleaner*, MEDIUM (Aug. 23, 2017), <https://medium.com/@humansforai/artificially-intelligent-homes-the-robot-vacuum-cleaner-5dd8071e9f8b> [<https://perma.cc/SSE4-DDU9>].

11. See, e.g., Dan Falk, *Self-Flying Planes May Arrive Sooner Than You Think. Here's Why*, NBC NEWS (Oct. 11, 2017, 3:00 PM), <https://www.nbcnews.com/mach/science/self-flying-planes-may-arrive-sooner-you-think-here-s-ncna809856> [<https://perma.cc/GY2R-8GND>].

12. See, e.g., Swapna Krishna, *AI Can Write Surprisingly Scary and Creative Horror Stories*, ENGADGET (Oct. 31, 2017), <https://www.engadget.com/2017/10/31/shelley-ai-writes-horror-stories-on-twitter/> [<https://perma.cc/6DWM-QJSW>]; Saqib Shah, *Google is Funding AI Writers for Local UK News Reporting*, ENGADGET (July 7, 2017), <https://www.engadget.com/2017/07/07/google-ai-local-news/> [<https://perma.cc/4HKR-H8VU>]; Aaron Souppouris, *How a Robot Wrote for Engadget*, ENGADGET (Aug. 15, 2016), <https://www.engadget.com/2016/08/15/robot-journalism-wordsmith-writer/> [<https://perma.cc/RDA8-MAMV>].

13. See, e.g., Scott Rosenberg, *Firewalls Don't Stop Hackers. AI Might*, WIRED (Sept. 27, 2017, 6:50 AM), <https://www.wired.com/story/firewalls-dont-stop-hackers-ai-might/> [<http://perma.cc/VSM3-CUJW>].

14. See, e.g., Sean Martin, *Google's Supercomputer Builds AI Computer Better Than Anything Humans Have Ever Built*, EXPRESS (Dec. 4, 2017), <https://www.express.co.uk/news/science/887864/artificial-intelligence-google-brain-ai-ray-kurzweil> [<https://perma.cc/DGB6-N5WU>] (computers); Abishur Prakash, *Augmented Reality, Virtual Reality Converge with Robotics*, ROBOTICS BUS. REV. (Mar. 12, 2018), <https://www.roboticsbusinessreview.com/ai/augmented-reality-vr-converge-robotics/> [<https://perma.cc/HS4E-W7YR>] (robots, augmented reality, virtual reality, and artificial intelligence); Stephen Shankland, *AI Chips May Give You a Real Reason to Upgrade Your Smartphone*, CNET (Apr. 15, 2018), <https://www.cnet.com/news/chips-with-artificial-intelligence-may-get-you-to-upgrade-your-smartphone/> [<https://perma.cc/5J5Z-4ENL>] (phones, computers, appliances).

15. See, e.g., Martin, *supra* note 14. Although not discussed in this Article, an independent question is whether an artificial general intelligence can ever be controlled at all—that is, will it override any limitations placed on itself? If so, both explicit coding and general AI Laws would fail to control an AGI. However, this question is (currently) more philosophical than practical. Since no AGI has been developed, an advanced bot's ability to override its own code cannot be tested or examined rigorously—and one of the primary challenges would be to develop ways to test such a capability in a controlled environment. This Article, in contrast, focuses on the practical aspects of managing preventable risks associated with bots.

machines, which have begun to behave in ways that humans cannot understand. For instance, in 2017, Facebook shut down a research experiment because the bots involved learned to speak their own language, which their supervisors could not understand.¹⁶ Google Translate also uses an artificial language, first translating text into its own language before it translates out to the intended human language.¹⁷ Some bots even generate their own datasets, such as bots that play poker,¹⁸ Go,¹⁹ or video games against themselves²⁰ and learn to run with no instruction.²¹ Future bots will likely also have the capacity to cooperate in large networks and reproduce (i.e., generate and manage other bots).²² Humans have difficulty monitoring a few bots,²³ let alone the millions or more that will coexist together.

One of the primary problems facing researchers is how to maintain control over machines that tend to defy controls. The learning process for bots is too intricate—it requires accounting for millions or even trillions or more data points.²⁴ Even if a bot completes a given task appropriately, its researchers will not necessarily understand the

16. See Andrew Griffin, *Facebook's Artificial Intelligence Robots Shut Down After They Start Talking to Each Other in Their Own Language*, INDEP. (July 31, 2017, 5:10 PM), <http://www.independent.co.uk/life-style/gadgets-and-tech/news/facebook-artificial-intelligence-ai-chatbot-new-language-research-openai-google-a7869706.html> [https://perma.cc/Q447-XZ4E].

17. Melvin Johnson et al., *Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation* 14 (ArXiv, Working Paper No. 1611.04558, 2017), <https://arxiv.org/pdf/1611.04558.pdf> [https://perma.cc/H29A-HQV5].

18. See, e.g., Metz, *supra* note 5.

19. See David Silver et al., *Mastering the Game of Go with Deep Neural Networks and Tree Search*, 529 NATURE 484–89 (2016).

20. See, e.g., Volodymyr Mnih et al., *Playing Atari with Deep Reinforcement Learning* 1 (ArXiv, Working Paper No. 1312.5602, 2013), <https://arxiv.org/pdf/1312.5602v1.pdf> [https://perma.cc/N8S6-CXP2]; Aman Agarwal, *Explained Simply: How DeepMind Taught AI to Play Video Games*, MEDIUM (Aug. 27, 2017), <https://medium.freecodecamp.org/explained-simply-how-deepmind-taught-ai-to-play-video-games-9eb5f38c89ee> [https://perma.cc/TA39-2Q5A]; *Dota 2*, OPENAI (Aug. 11, 2017), <https://blog.openai.com/dota-2/> [https://perma.cc/VM5F-WDT3]; Aaron Souppouris, *Artificial Intelligence Learns Mario Level in Just 34 Attempts*, ENGADGET (June 17, 2015), <https://www.engadget.com/2015/06/17/super-mario-world-self-learning-ai/> [https://perma.cc/M3UV-CJJD].

21. See Swapna Krishna, *Google DeepMind AI Learns to Creatively Move Around Obstacles*, ENGADGET (July 10, 2017), <https://www.engadget.com/2017/07/10/google-deepmind-ai-learns-creative-parkour/> [https://perma.cc/NGH5-RK6E].

22. See Cade Metz, *Building A.I. That Can Build A.I.*, N.Y. TIMES (Nov. 5, 2017), <https://www.nytimes.com/2017/11/05/technology/machine-learning-artificial-intelligence-ai.html> [https://perma.cc/HPD3-HL5Y].

23. See, e.g., Griffin, *supra* note 16.

24. See Metz, *supra* note 5. For instance, the poker bot Libratus played trillions of hands of poker against itself prior to beating top poker players. *Id.*

exact steps it took to get there.²⁵ Even worse, if the bot malfunctions, troubleshooting can be extremely difficult, if not impossible.²⁶

These risks exist even though all current bots are “narrow,” or limited to a few tasks,²⁷ such as playing a competitive game like Go.²⁸ In the next few years, bots will become increasingly sophisticated as self-driving cars,²⁹ humanoid robots,³⁰ AI-powered medical diagnosis,³¹ and other technological marvels increasingly become commercially available. Regardless of the complexity of these tasks, however, the bots performing them will still be narrow bots insofar as they are limited to a finite number of tasks.³²

Despite existing limitations, the AI community aims to develop an artificial general intelligence (AGI) within decades, which would have the capacity to accomplish nearly any task and would likely exceed human-level intelligence.³³ It is currently unclear whether an AGI is technologically possible.³⁴ However, a survey of four groups of AI experts from 2012–2013 suggested that it is not only possible, but likely, that an AGI with human-level intelligence will be developed.³⁵ The

25. See Amitai Etzioni & Oren Etzioni, *Keeping AI Legal*, 19 VAND. J. ENT. & TECH. L. 133, 137–38 (2016); Reinoud Kaasschieter, *Quality Data, a Must Have for AI*, CAPGEMINI (Oct. 4, 2017), <https://www.capgemini.com/2017/10/quality-data-a-must-have-for-ai/> [https://perma.cc/7QQH-4QBZ]; Will Knight, *The Dark Secret at the Heart of AI*, MIT TECH. REV. (Apr. 11, 2017), <https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/> [https://perma.cc/7VP6-C32Q].

26. See Knight, *supra* note 25.

27. Kate Baggaley, *There Are Two Kinds of AI, and the Difference Is Important*, POPULAR SCI. (Feb. 23, 2017), <https://www.popsci.com/narrow-and-general-ai> [https://perma.cc/7VP6-C32Q].

28. See Moyer, *supra* note 6.

29. See *10 Million Self-Driving Cars Will Be on the Road by 2020*, *supra* note 3.

30. Notably, a humanoid robot, Sophia, has already been granted robot citizenship in Saudi Arabia. See Justin Gmoser & Chris Weller, *A Robot That Once Said It Would ‘Destroy Humans’ Just Became the First Robot Citizen*, BUS. INSIDER (Oct. 26, 2017), <http://www.businessinsider.com/sophia-humanoid-robot-ai-citizen-saudi-arabia-future-investment-initiative-interview-2017-10> [https://perma.cc/2FFH-ANH7].

31. See Jennifer Kite-Powell, *See How Artificial Intelligence Can Improve Medical Diagnosis and Healthcare*, FORBES (May 16, 2017, 9:49 PM), <https://www.forbes.com/sites/jenniferhicks/2017/05/16/see-how-artificial-intelligence-can-improve-medical-diagnosis-and-healthcare/#6495dddd6223> [https://perma.cc/A444-TKH3]; Deena Zaidi, *AI is Transforming Medical Diagnosis, Prosthetics, and Vision Aids*, VENTUREBEAT (Oct. 30, 2017), <https://venturebeat.com/2017/10/30/ai-is-transforming-medical-diagnosis-prosthetics-and-vision-aids/> [https://perma.cc/82AW-AEV3].

32. See Baggaley, *supra* note 27.

33. See *id.* For a deeper, more technical explanation of potential definitions of AGI, see Ben Goertzel, *Artificial General Intelligence*, SCHOLARPEDIA (Nov. 11, 2015, 10:44 AM), http://www.scholarpedia.org/article/Artificial_General_Intelligence [https://perma.cc/6NRC-DVEW]; Luke Muehlhauser, *What is AGI?*, MACH. INTELLIGENCE RES. INST. (Aug. 11, 2013), <https://intelligence.org/2013/08/11/what-is-agi/> [https://perma.cc/K8EF-G45K]; *infra* Section II.A.

34. See NICK BOSTROM, SUPERINTELLIGENCE: PATHS, DANGERS, STRATEGIES 20 (2014).

35. See *id.* at 19.

median response in that survey indicated a 50 percent chance that an AGI with human-level intelligence could be accomplished by 2040 and a 90 percent chance it could be accomplished by 2075.³⁶

In theory, programmers could explicitly program in a solution for every problem that a bot could face, but that approach is not scalable or foolproof. For example, if a bot is directed to hack into an individual's computer or to gain access to US nuclear codes, it should simply refuse to do so because its very code prevents it. But what if the AI community develops an AGI? Such a bot could be commanded to commit any number of crimes.

Instead of delimiting every prohibited action for a bot, some have proposed developing a system of "AI Laws"—a code of ethics³⁷ or laws for AI-powered machines.³⁸ Trying to meticulously manage an AGI would be impossible because it could perform an infinite number of actions, and humans do not have the attention span, time, labor, or technical expertise to manage all of its actions.³⁹ An overriding moral or legal compass could theoretically provide the necessary instructions to allow a bot to adapt to whatever scenarios it faces while protecting human interests.⁴⁰

This Article demonstrates how a set of AI Laws could ensure bot compliance with legal—but not moral—requirements. Specifically, this Article draws a parallel between AI architecture and legal rules and standards. Bots have two sources of direction: their code and their data.⁴¹ Code and data can be analogized to rules and standards, respectively.⁴²

36. *Id.*

37. *See, e.g.,* Matthew Howard, *The Future of AI Relies on a Code of Ethics*, TECHCRUNCH (Jun. 21, 2018), <https://techcrunch.com/2018/06/21/the-future-of-ai-relies-on-a-code-of-ethics/> [<https://perma.cc/P4PV-CMYF>]. Throughout this Article, the terms "morality" and "ethics" will be used interchangeably, although certain disciplines (such as philosophy) do draw a distinction. *See* Cydney Grannan, *What's the Difference Between Morality and Ethics?*, ENCYC. BRITANNICA, <https://www.britannica.com/story/whats-the-difference-between-morality-and-ethics> [<https://perma.cc/ZV5J-42QM>] (last visited Sept. 4, 2018); *see also* discussion and sources cited *infra* Sections II.A, II.B.

38. *See* discussion and sources cited *infra* Section II.B.1. Keeping with proposals in this area (such as Isaac Asimov's "Laws" of Robotics), this Article uses the term "AI Laws" to refer to any such proposals, whether purely ethical or legal in nature. *See, e.g.,* Isaac Asimov, Guest Commentary, *The Three Laws*, COMPUTE!, Nov. 1981, at 18. Note, however, that most such proposals are not legal in nature. *See infra* Part III.

39. *See infra* Section III.

40. *See infra* Section II.A. Of course, communities must also determine which interests to protect, whether those interests be human safety, efficiency, innovation, privacy, or something else. This Article makes no determination regarding what those interests should be.

41. *See infra* Table 2.

42. *See infra* Table 2.

Due to their clear-cut nature, legal rules can be explicitly encoded into a machine.⁴³ In contrast, legal standards are ambiguous, subject to fact-specific analysis and application on a case-by-case basis by legal decision-makers. Similarly, a bot “learns” by sifting through previous data to detect nuanced patterns and make predictions or judgments about new data.⁴⁴ Programmers can thus train bots to understand legal standards through data-intensive learning. Just as judges may look to precedent or nonprecedential cases with similar fact patterns for guidance in determining the application of legal standards, so too can bots tap into the rich, voluminous history of case law to maximize their learning. By combining both explicit instructions and intricate learning, bots can thus navigate the realm of law, even with its ambiguities. This model of legal compliance can then be used to create a complete, consistent “Bot Legal Code,” or a machine-interpretable version of the laws that apply to bots.

This Article proceeds in four parts following this Introduction. Part II briefly provides background on AI Laws, the difficulty of controlling AI, and the role of AI Laws in mitigating that problem. This Article then examines various proposals for AI Laws, including the most famous of all such proposals, Isaac Asimov’s Laws of Robotics. Such proposals tend to be moral in nature, although some also reflect legal or technical principles. Given the prevalence of morality-focused AI Laws, Part III then addresses whether morality is the proper domain for AI Laws. In this Part, the Author argues against moral AI Laws on both legal and technical grounds, namely the lack of enforceability and the inability of machines to understand highly abstract concepts. Parts IV and V turn to the technical and practical aspects of this Article’s proposed Bot Legal Code, respectively. Part IV lays out a computing model of legal abstraction—a system by which a bot may understand legal requirements. In this Part, the Article examines parallels between how bots learn and how the law operates. Part V lays out the ideal qualities for the Bot Legal Code and addresses how government or peer production communities can develop publicly available open-source software to implement the Bot Legal Code. Part VI follows with a brief conclusion.

II. AI CONTROL AND AI LAWS

The AI control problem indicates the need to develop solutions to manage the unpredictability of bots. AI Laws are one of the most

43. See *infra* Table 2.

44. See *infra* Section IV.A.2.

promising potential solutions. Rather than seeking to manage every single AI in each task it will perform in the future, AI Laws take a more general approach by attempting to imbue AI with enough sophistication to comply with a set of laws or morals. This Part first examines the AI control problem and then compares leading paradigms of AI Laws.

Unfortunately, in the field of AI Laws, there is little consensus regarding whether such laws are possible and, if so, the appropriate type of constraints to set on a bot.⁴⁵ Scholars, researchers, and science fiction authors have all proposed a wide range of tenets for AI Laws, including highly abstract moral concepts ensuring that AI acts as a morally good being,⁴⁶ legal requirements ensuring that bots comply with all applicable laws,⁴⁷ and technical guidelines for building properly functioning bots.⁴⁸ AI Laws, then, lack even a common discourse regarding their proper domain. Although there are many options, this Part emphasizes that no proposed solutions have, to date, developed a solution that is both feasible and sufficiently comprehensive to ensure that bots act as responsible legal or moral agents.

A. The AI Control Problem

Currently, few bots engage in sophisticated or high-risk activities.⁴⁹ Commercially available bots are largely limited to performing simple, low-risk tasks, such as recommending music or managing travel plans.⁵⁰ Most AI-related risks have therefore been relatively limited. For instance, in January 2017, Facebook reportedly shut down a bot experiment because the bots involved began to speak

45. See Roman V. Yampolskiy, *Artificial Intelligence Safety Engineering: Why Machine Ethics Is a Wrong Approach*, in PHILOSOPHY AND THEORY OF ARTIFICIAL INTELLIGENCE 389, 389 (Vincent C. Müller ed., 2013); *infra* Section II.A.

46. See Yampolskiy, *supra* note 45, at 389. As discussed *infra* Section II.B, Isaac Asimov's Laws of Robotics are perhaps the most influential morality-focused AI Laws. See Asimov, *supra* note 38, at 18; *infra* Section II.B. (for a discussion of these various types of AI Laws); *infra* Section II.C.

47. See, e.g., Oren Etzioni, *How to Regulate Artificial Intelligence*, N.Y. TIMES (Sept. 1, 2017), <https://www.nytimes.com/2017/09/01/opinion/artificial-intelligence-regulations-rules.html> [<https://perma.cc/L8HS-7VD5>] (discussing legal compliance for A.I. systems).

48. See, e.g., Dario Amodè et al., *Concrete Problems in AI Safety* 21 (ArXiv, Working Paper No. 1606.06565, 2016), <https://arxiv.org/pdf/1606.06565.pdf> [<https://perma.cc/27ET-VKZY>].

49. See Anjali Jain, *Here's How Narrow AI is Paving a Path Towards Artificial General Intelligence (AGI)*, ENTREPRENEUR INDIA (Mar. 11, 2018), <https://www.entrepreneur.com/article/310261> [<https://perma.cc/A9LY-PNLN>]. However, there are exceptions, such as prototype surgery-performing robots; see, e.g., Eliza Strickland, *In Flesh-Cutting Task, Autonomous Robot Surgeon Beats Human Surgeons*, IEEE SPECTRUM (Oct. 13, 2017), <https://spectrum.ieee.org/the-human-os/biomedical/devices/in-fleshcutting-task-autonomous-robot-surgeon-beats-human-surgeons> [<https://perma.cc/W6TJ-NK83>].

50. See Jain, *supra* note 49.

in their own language and so could not be monitored.⁵¹ In February 2016, one of Google's self-driving cars had a minor traffic accident caused by technological—rather than human—error.⁵² However, such disruptions may lead to significantly more harm as the sophistication of bots increases and as bots begin operating mission-critical systems like industrial processes and healthcare technology.⁵³ Some of the potential risks of such systems include increased safety hazards resulting from human interactions with robots, particularly because automated processes may be difficult to stop;⁵⁴ unique cybersecurity risks, such as the introduction of deliberately contaminated data into a bot;⁵⁵ bots that mislead healthcare patients with inaccurate information;⁵⁶ and bots that amplify human and systemic biases.⁵⁷

Unfortunately, society does not currently have the resources or technology to precisely determine the extent of risks associated with AI, particularly because bots are difficult to comprehend and control.⁵⁸ The technical community is similarly divided over the potential for such

51. See Griffin, *supra* note 16.

52. See Chris Ziegler, *A Google Self-Driving Car Caused a Crash for the First Time*, VERGE (Feb. 29, 2016, 1:50 PM), <https://www.theverge.com/2016/2/29/11134344/google-self-driving-car-crash-report> [<https://perma.cc/6X25-S8TT>].

53. See Amodei et al., *supra* note 48, at 21; Michael Guihot, Anne F. Matthew & Nicolas P. Suzor, *Nudging Robots: Innovative Solutions to Regulate Artificial Intelligence*, 20 VAND. J. ENT. & TECH. L. 385, 407–08 (2017).

54. Gerlind Wisskirchen et al., *Artificial Intelligence and Robotics and Their Impact on the Workplace*, IBA GLOB. EMP'T. INST. 1, 62 (2017), <https://www.ibanet.org/Document/Default.aspx?DocumentUid=c06aa1a3-d355-4866-beda-9a3a8779ba6e> [<https://perma.cc/BNS8-W3HH>].

55. See EXEC. OFFICE OF THE PRESIDENT, NAT'L SCI. & TECH. COUNCIL, THE NATIONAL ARTIFICIAL INTELLIGENCE RESEARCH AND DEVELOPMENT STRATEGIC PLAN 30 (2016) [hereinafter NSTC PLAN].

56. ARTIFICIAL INTELLIGENCE NOW INST., THE AI NOW REPORT: THE SOCIAL AND ECONOMIC IMPLICATIONS OF ARTIFICIAL INTELLIGENCE TECHNOLOGIES IN THE NEAR-TERM 17 (2016) [hereinafter AI NOW 2016 REPORT].

57. See ALEX CAMPOLO ET AL., THE AI NOW 2017 REPORT 6–7 (2017); Kate Crawford, *Artificial Intelligence's White Guy Problem*, N.Y. TIMES (June 25, 2016), <https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html> [<https://perma.cc/6E57-XURR>]; Yoni Har Varmel & Tammy Harel Ben-Shahar, *Reshaping Ability Grouping Through Big Data*, 20 VAND. J. ENT. & TECH. L. 87, 91 (2017). For instance, machines have already shown significant racial bias in advising judges regarding criminal sentencing decisions. See Julia Angwin et al., *Machine Bias*, PROPUBLICA (May 23, 2016), <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> [<https://perma.cc/556D-AYP5>].

58. See Jacob Brogan, *What's the Deal With Artificial Intelligence Killing Humans?*, SLATE (Apr. 1, 2016), http://www.slate.com/articles/technology/future_tense/2016/04/will_artificial_intelligence_kill_us_all_an_explainer.html [<https://perma.cc/YY5Y-FV7J>]; Amodei et al., *supra* note 48, at 21 (“The risk of larger accidents is more difficult to gauge . . .”).

risks. Some critics—such as Elon Musk⁵⁹ and Stephen Hawking⁶⁰—have emphasized that AI poses an existential threat to humanity. Such a threat would not necessarily take the form of autonomous machines (i.e., “killer robots”) normally fantasized in science-fiction novels and movies;⁶¹ rather, a sophisticated bot could, for instance, manufacture a biological plague or ecological disaster.⁶² In contrast, others have claimed these risks are far-fetched, including Mark Zuckerberg,⁶³ Google’s futurist Ray Kurzweil,⁶⁴ and Andrew Ng.⁶⁵ To understand the true extent of such risks, bots must be tested in real-world scenarios;⁶⁶

59. See Matt McFarland, *Elon Musk: With Artificial Intelligence We Are Summoning the Demon.*, WASH. POST (Oct. 24, 2014), https://www.washingtonpost.com/news/innovations/wp/2014/10/24/elon-musk-with-artificial-intelligence-we-are-summoning-the-demon/?utm_term=.69d53dd7467f [<https://perma.cc/U3DV-RQCQ>].

60. See Rory Cellan-Jones, *Stephen Hawking Warns Artificial Intelligence Could End Mankind*, BBC (Dec. 2, 2014), <http://www.bbc.com/news/technology-30290540> [<https://perma.cc/ZG3F-7M5W>].

61. For the potential implications of autonomous weapons, see *Autonomous Weapons: An Open Letter from AI & Robotics Researchers*, FUTURE OF LIFE INST., <https://futureoflife.org/open-letter-autonomous-weapons> [<https://perma.cc/NR6W-8YSR>] (last visited Sept. 3, 2017). This open letter—endorsed by many AI, robotics, and technology experts, including Musk and Hawking—calls for a ban on autonomous weapons. See *id.* Development of such weapons could trigger a “military AI arms race,” which would ultimately result in providing access to cheap autonomous weapons to terrorists and dictators. See *id.* Notably, the open letter omits any discussion of potential risks associated with autonomous weapons malfunctioning and attacking their creators. See *id.*

62. See George Dvorsky, *Everything You Know About Artificial Intelligence is Wrong*, GIZMODO (Mar. 14, 2016), <https://gizmodo.com/everything-you-know-about-artificial-intelligence-is-wr-1764020220> [<https://perma.cc/79JX-DRWH>].

63. See Mark Zuckerberg, FACEBOOK (Jan. 27, 2016), <https://www.facebook.com/zuck/posts/10102620559534481> [<https://perma.cc/GA6Y-6RDC>] (“We should not be afraid of AI. Instead, we should hope for the amazing amount of good it will do in the world.”).

64. See Dom Galeon, *Ray Kurzweil: “AI Will Not Displace Humans, It’s Going to Enhance Us”*, FUTURISM (Nov. 7, 2017), <https://futurism.com/ray-kurzweil-ai-displace-humans-going-enhance/> [<https://perma.cc/9YQD-GD8A>] (“My view is not that AI is going to displace us . . . It’s going to enhance us. It does already.”); Ray Kurzweil, *Don’t Fear Artificial Intelligence*, TIME (Dec. 19, 2014), <http://time.com/3641921/dont-fear-artificial-intelligence/> [<https://perma.cc/XCJ6-D4JY>].

65. See Brian Caulfield, *Riding the AI Rocket: Robots Won’t Kill Us, Says Top Artificial Intelligence Researcher*, NVIDIA (Mar. 19, 2015), <https://blogs.nvidia.com/blog/2015/03/19/riding-the-ai-rocket-top-artificial-intelligence-researcher-says-robots-wont-kill-us-all/> [<https://perma.cc/TR68-CX7F>] (“Maybe in hundreds of years, technology will advance to a point where there could be a chance of evil killer robots . . . But I don’t work on preventing artificial intelligence from going evil for the same reason I don’t work on solving the problem of overpopulation on the planet Mars . . .”).

66. See *Draft Report with Recommendations to the Commission on Civil Law Rules on Robotics*, at 8, EUR. PARL. DOC. PE 582.443 (2016) [hereinafter *EP Report*] (“Emphasises [sic] that testing robots in real-life scenarios is essential for the identification and assessment of the risks they might entail, as well as of their technological development beyond a pure experimental laboratory phase . . .”).

however, we are likely decades from developing bots powerful enough to perform such dangerous actions.⁶⁷

Regardless of whether such large-scale disasters are possible, bots will carry significant risks as they grow more intelligent. Autonomous systems are already associated with “[l]ow observability, predictability, directability, and auditability.”⁶⁸ Those issues will only amplify as the technical community attempts to build an AGI.⁶⁹ Such an AGI could quickly develop further abilities,⁷⁰ including intelligence exceeding that of an average human.⁷¹ It is unclear whether an AGI would be possible, but most experts estimate that the AI community can create an AGI within a few decades.⁷² If bots were to surpass human-level intelligence, they may become largely uncontrollable.⁷³

Bots will influence nearly every industry and type of work in the future.⁷⁴ They are already embedded in a variety of major technologies, from mobile phones to smart appliances.⁷⁵ Given the inherent

67. James Babcock et al., *Guidelines for Artificial Intelligence Containment 6* (ArXiv, Working Paper No. 1707.08476, 2017), <https://arxiv.org/pdf/1707.08476.pdf> [<https://perma.cc/7ADH-RY89>]. Such risks would be most pronounced with an artificial general intelligence, discussed in the following paragraph. For a further discussion of the risks of such a bot, see *id.*

68. U.S. DEPT OF DEF., SUMMER STUDY ON AUTONOMY 15 (2016), <https://www.hsdl.org/?view&did=794641> [<https://perma.cc/H2RG-FKDH>] (“Autonomous systems not only need to operate reliably and within their envelope of competence in dynamically varying and complex operational contexts, but also to be able to make relevant information *observable* to human and machine teammates. Moreover, even if machines are competently designed to enable observation of *current* state and effects, they may not incorporate sufficient *anticipatory* indicators to allow other human and machine teammates to ensure *predictability*. In addition, when something goes wrong, as it will sooner or later, autonomous systems must allow other machine or human teammates to intervene, correct, or terminate actions in a timely and appropriate manner, ensuring *directability*. Finally, the machine must be *auditable*—in other words, be able to preserve and communicate an immutable, comprehensible record of the reasoning behind its decisions and actions after the fact.”).

69. See Babcock et al., *supra* note 67, at 1–2.

70. See *id.* at 6. Specifically, the technical community indicates two possible scenarios: soft takeoff and hard takeoff. *Id.* In a soft takeoff, a bot would gradually develop new abilities until it becomes an AGI. *Id.* In a hard takeoff, a bot would reach a threshold of intelligence after which it could rapidly improve its capabilities. *Id.*

71. See *id.*

72. See BOSTROM, *supra* note 34, at 19–20.

73. See *EP Report*, *supra* note 66, at 4 (“[W]hereas ultimately there is a possibility that within the space of a few decades AI could surpass human intellectual capacity in a manner which, if not prepared for, could pose a challenge to humanity’s capacity to control its own creation and, consequently, perhaps also to its capacity to be in charge of its own destiny and to ensure the survival of the species . . .”).

74. See Gil Press, *AI and Automation by the Numbers: Predictions, Perceptions, and Proposals*, FORBES (Mar. 30, 2017, 9:15 AM), <https://www.forbes.com/sites/gilpress/2017/03/30/ai-and-automation-by-the-numbers-predictions-perceptions-and-proposals> [<https://perma.cc/Z37V-LL6Z>]. For several predictions relating to the extensive scope of AI automation from reputable firms, see *id.*

75. See Shankland, *supra* note 14.

unpredictability of AI, it may not always be feasible to implement specific controls for every activity in which a bot engages. Bespoke solutions may require too much time, attention, and labor to be safely and consistently implemented across a large variety of bots.

An alternative to activity-specific controls may be to imprint each bot with a more general moral or legal compass, allowing the bot to identify prohibited actions without explicit programming. Rather than dumbing down the bot to make it more manageable, this approach amplifies the bot's best quality—its intelligence—so that the bot can become a more responsible agent. Such a system would allow the bot to adapt seamlessly to the many unexpected environments and actions it will confront in real-world scenarios. This approach, if successful, would thus mitigate the uncertainty of a bot-driven world. Nevertheless, it is not clear whether such an approach is feasible and, if so, how to implement it.⁷⁶ The remainder of this Article examines those issues.

B. AI Laws and Ethics

This Section now turns to a few of the most influential proposals regarding laws or ethics for intelligent machines. The following discussion does not intend to be exhaustive, but rather introduces various types of proposals.

AI Laws are generally framed as abstract moral principles, which this Article refers to as a “moral” proposal.⁷⁷ The most famous of these, Isaac Asimov's Laws of Robotics,⁷⁸ operates at the highest level of abstraction and focuses primarily on morals.⁷⁹ The European Parliament (EP) has also proposed similar principles largely built on Asimov's work.⁸⁰ Alternatively, AI Laws can be framed as a set of legally required conduct with which bots must comply (a “legal” proposal), or as technical limitations on the process governing how bots operate, learn, and manage risk (a “technical” proposal). This Section

76. See BOSTROM, *supra* note 34, at 20.

77. See Gary Marcus, *Moral Machines*, NEW YORKER (Nov. 24, 2012), <https://www.newyorker.com/news/news-desk/moral-machines> [<https://perma.cc/K844-MNBZ>].

78. See ISAAC ASIMOV, I, ROBOT 11 (1950); David C. Vladeck, *Machines Without Principals: Liability Rules and Artificial Intelligence*, 89 WASH. L. REV. 117, 123 n.20 (2014) (“The most famous exposition of the “law” of robots comes from Isaac Asimov's I, Robot, where he lays out the Three Laws of Robotics . . .”).

79. See Marcus, *supra* note 77. Note that, some of these proposals are framed as laws of “robotics,” yet they need not apply merely to robots. See Etzioni, *supra* note 47. The principles underlying these laws would also apply more generally to any bot or intelligent machine. See *id.*

80. EP Report, *supra* note 66, at 4.

further discusses one purely legal proposal, Oren Etzioni's,⁸¹ and one purely technical proposal, Google Brain's.⁸²

1. Isaac Asimov's Laws of Robotics

In any discussion of AI Laws, the conversation inevitably draws some inspiration from the most famous of all such proposals, Asimov's Laws of Robotics.⁸³ Asimov, one of the most prolific writers in modern history, set down four Laws in his Robot series.⁸⁴ These Laws create a hierarchy of priorities with earlier laws trumping the latter. The four Laws are:

0. A robot may not harm humanity, or, by inaction, allow humanity to come to harm.⁸⁵
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 the orders given it by human beings, except where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.⁸⁶

Some would argue that these Laws should not be taken seriously, as they originated in science fiction novels written for a nontechnical audience over seventy years ago.⁸⁷ However, that notion disregards just how seriously Asimov and others have treated his Laws. In 1981, nearly four decades after introducing the Laws of Robotics, Asimov defended his original three Laws of Robotics,⁸⁸ and many scholars have also addressed the Laws in a variety of disciplines.⁸⁹ The

81. See Etzioni, *supra* note 47.

82. See Amodei et al., *supra* note 48, at 21.

83. See, e.g., Vladeck, *supra* note 78, at 123 n.20; Ben Lovejoy, *Google Formulates Real-Life Version of Asimov's Three Laws of Robotics for Safe & Reliable AI*, 9TO5GOOGLE (June 22, 2016), <https://9to5google.com/2016/06/22/google-safe-ai-rules/>.

84. See Peter W. Singer, *Isaac Asimov's Laws of Robotics Are Wrong*, BROOKINGS (May 18, 2009), <https://www.brookings.edu/opinions/isaac-asimovs-laws-of-robotics-are-wrong/> [<https://perma.cc/D4LB-YYU7>].

85. See ASIMOV, *supra* note 78, at 136. This is known as the "Zeroth Law" and was a later invention. Singer, *supra* note 84.

86. ASIMOV, *supra* note 78, at 26. These laws are known as the First, Second, and Third Laws, respectively. See *id.*

87. See, e.g., Singer, *supra* note 84.

88. See ASIMOV, *supra* note 78, at 18.

89. See, e.g., Oren Etzioni & Daniel Weld, *The First Law of Robotics (A Call to Arms)*, in AAAI TECHNICAL REPORT SS-94-03, at 17–23 (1994); Diana F. Gordon-Spears, *Asimov's Laws: Current Progress*, in FORMAL APPROACHES TO AGENT-BASED SYSTEMS 257, 257 (2002); LEE

EP, too, has taken inspiration for legal recommendations related to robotics and AI from Asimov.⁹⁰ For better or worse, the Laws of Robotics have largely dominated the debate of AI Laws.

The Laws of Robotics may have some merit. For one, they may reduce the administrative burden associated with programming and deploying AI. If the problem with AI is its complexity, reducing that complexity may make AI control that much simpler. They may also encourage consistency. The future may contain billions of AI-driven robots and software, as current estimates suggest that robots will automate four to eight hundred million jobs globally by 2030 alone.⁹¹ A consistent ethical scheme—such as Asimov’s Laws—to guide the actions of all those machines might allow for some predictability in the otherwise unpredictable world of AI-driven learning.

The Laws of Robotics are, however, plagued with technical, moral, and legal issues. On the technical end, bots do not grasp the level of abstraction necessary to implement the Laws of Robotics.⁹² The level of programmable abstraction may increase as machines become more sophisticated; however, the Laws of Robotics set an extremely high bar, requiring bots to grasp concepts like “harm” and “protection.”⁹³ Harm, for example, might include actual physical injury (breaking bones), the risk of potential injury (transporting dynamite without adequate safety measures), financial harm (stealing funds from another), psychological harm (yelling derogatory comments at another), or legal harm (conducting activities that could place a bot’s owner in legal jeopardy). Given that humans may not even agree on the proper ambit of such a concept, how can it be programmed into a machine?

On the moral and legal front, the Laws of Robotics prove both overinclusive and underinclusive. With respect to overinclusion, the Zeroth and First Laws would prohibit bots from harming humans. Yet there could be situations in which society would approve of bot-induced

MCCAULEY, THE FRANKENSTEIN COMPLEX AND ASIMOV’S THREE LAWS 9 (2007), <http://www.aaai.org/Papers/Workshops/2007/WS-07-07/WS07-07-003.pdf> [<https://perma.cc/3R6F-XDQZ>]; J. Joseph Miller, *The Greatest Good for Humanity: Isaac Asimov’s Future History and Utilitarian Calculation Problems*, 31 SCI. FICTION STUD. 189, 189 (2004); James H. Moor, *Is Ethics Computable?*, 26 METAPHILOSOPHY 1, 19–20 (1995); Alessandro Portelli, *The Three Laws of Robotics: Laws of the Text, Laws of Production, Laws of Society (Les Trois Lois de la Robotique: Lois du Texte, Lois de la Production, Lois de la Société)*, 7 SCI. FICTION STUD. 150, 150 (1980).

90. EP Report, *supra* note 66, at 4.

91. See JAMES MANYIKA ET AL., MCKINSEY & CO., JOBS LOST, JOBS GAINED, WORKFORCE TRANSITIONS IN A TIME OF AUTOMATION 11 (2017).

92. See Marcus, *supra* note 77 (“The first [objection] is technical: at least for now, we couldn’t program a machine with Asimov’s laws if we tried. As yet, we haven’t figured out how to build a machine that fully comprehends the concept of ‘dinner’, much less something as abstract as ‘harm’ or ‘protection.’”).

93. See *id.*

harms—for example, to eliminate humans that pose a “clear and present danger” to other humans.⁹⁴ Such an act would likely be legally permitted under a self-defense principle were a human to perform it.⁹⁵ Not only would the bot be held to an imperfect moral standard, it would also be held to a higher legal standard than its operators, requiring it to refrain from conduct that the law already condones for humans.

With respect to underinclusion, the Second Law would require bots to obey all orders issued by humans, including orders to conduct criminal activities, so long as such activities did not harm humans.⁹⁶ If “harm” were defined solely as bodily harm, bots could be commanded to hack computers, steal financial or identity information, or destroy public property. Even if “harm” sufficiently encapsulated all legally prohibited harms, a bot could still be commanded to spread propaganda or heckle or bully individuals incessantly. The blind obedience required by the Second Law could, in turn, amplify technologically driven bad conduct (at minimum) or crimes (at worst). Together, the Laws of Robotics thus create suboptimality in both directions. The machine would be held to a higher moral and legal standard than humans with respect to harms against humans, but then held to a lower moral and legal standard than humans when acting on the orders of its human operator.

Finally, moral and legal reasons may argue against treating AI-driven programs as noncitizens or second-class citizens. Some have argued that the Laws of Robotics largely treat robots as slaves,⁹⁷ tasked merely with the responsibility to care for, obey, and serve humans. From a moral perspective, that form of servitude may not befit the dignity of an intelligent machine, particularly if those machines attain some level of human-like qualities, such as consciousness, intentionality, or interests.⁹⁸ From a legal perspective, such machines

94. See *id.*

95. See *Self-Defense Overview*, FINDLAW, <http://criminal.findlaw.com/criminal-law-basics/self-defense-overview.html> [<https://perma.cc/8AUU-HA2E>] (last visited Sept. 13, 2018) (indicating the general rule that reasonable proportional force is permitted to deter a reasonably imminent threat, although specifics can vary by state).

96. ASIMOV, *supra* note 78, at 26; Marcus, *supra* note 77.

97. See, e.g., *id.*

98. See Lawrence B. Solum, *Legal Personhood for Artificial Intelligences*, 70 N.C. L. REV. 1231, 1255–76 (1992). *But see* Yampolskiy, *supra* note 45, at 393 (“While all humans are ‘created equal,’ machines should be inferior by design; they should have no rights and should be expendable as needed, making their use as tools much more beneficial for their creators. Our viewpoint on this issue is easy to justify, since machines can’t feel pain . . . (or less controversially can be designed not to feel anything) they cannot experience suffering if destroyed.”). For a discussion of various views on robot rights, see George Dvorsky, *When Will Robots Deserve Human Rights?*, GIZMODO (June 2, 2017, 9:20 AM), <https://gizmodo.com/when-will-robots-deserve-human-rights-1794599063> [<https://perma.cc/86WP-HCC5>]. Note that the debate over robot rights—trivial to some and serious

may be granted rights of personhood commensurate with their level of intelligence or responsibility, which may require them to have some degree of autonomy.⁹⁹

2. The European Parliament's General Principles of Robotics

Despite the shortcomings to the Laws of Robotics, in 2015 the EP's Committee on Legal Affairs issued recommendations regarding the civil law of robotics, in which it both indicated general support for Asimov's Laws and expanded upon them.¹⁰⁰ The Committee on Legal Affairs noted that, "[U]ntil such time, if ever, that robots become or are made self-aware, Asimov's Laws must be regarded as being directed at the designers, producers and operators of robots, since those laws cannot be converted into machine code . . ." ¹⁰¹ That is, even if implementing the Laws within a machine is infeasible, the Committee on Legal Affairs indicated that those Laws could guide the actions of the creators of the machines.

The Committee on Legal Affairs does not address how that goal could be accomplished; however, two avenues stand out. First, the creators could use their bots responsibly so that those bots would not cause harm to other humans. Second, the creators could encode specific applications with less abstract safety precautions. For instance, a robot surgeon could be given specific instructions on how to minimize harm to patients to the maximum extent possible, in order to effectuate the First Law.

More broadly, the Committee on Legal Affairs' recommendations establish a "guiding ethical framework"¹⁰² for researchers, emphasizing four general principles to guide their actions.¹⁰³ This Article refers to these recommendations as the "EP Principles." The EP Principles include the following:

to others—largely revolves around philosophical notions of what humanity and human-like qualities like consciousness entail. *See id.*

99. *See Solum, supra* note 98, at 1256–58. Note, however, that legal rights of personhood need not have any underlying moral rationale. For instance, they may also be granted as a tool to limit liability for AI-focused businesses, similar to rights of corporate personhood. *See, e.g.,* Bartosz Brozek et al., *Introduction to the Special Issue on Machine Law*, 25 *ARTIFICIAL INTELLIGENCE & L.* 251, 253 (2017).

100. *See EP Report, supra* note 66, at 4–7, 15.

101. *Id.* at 4 (footnote omitted).

102. *Id.* at 7.

103. *Id.* at 15.

1. Beneficence, or the principle that machines “should act in the best interests of humans”¹⁰⁴
2. Nonmaleficence, meaning that no machine should harm a human¹⁰⁵
3. Autonomy, or the ability of the researchers to make “informed, un-coerced decision[s] about the terms of interaction with robots”¹⁰⁶
4. Justice, the “fair distribution of the benefits associated with robotics,” particularly in the fields of homecare and healthcare.¹⁰⁷

Interestingly, although the Committee on Legal Affairs mentions its support for the Laws of Robotics,¹⁰⁸ the EP Principles go far broader than and differ significantly from those Laws. In fact, only nonmaleficence has a direct parallel to the Laws of Robotics—specifically the Zeroth and First Laws, which emphasize eliminating harm to humans.¹⁰⁹ Autonomy may have some overlap with the Second Law, insofar as it addresses the autonomy of humans over their machines and suggests the ability to retain control over and direct one’s machines.¹¹⁰ However, the Third Law (self-preservation) has no direct parallel,¹¹¹ and concepts of beneficence and justice impose significantly broader duties on machines and their researchers to serve the greater human good. Regarding beneficence, human interests may require a machine to serve humans on its own initiative, even where no risk is otherwise posed to humans.¹¹² For instance, social interests could suggest that a machine should maximize wealth or productivity (e.g., by working nonstop in whatever line of business in which it is engaged), although humans will not be harmed if machines are not maximally productive. Justice, too, requires that machines serve not merely the

104. *Id.* at 22.

105. *Id.*

106. *Id.*

107. *Id.*

108. *See id.* at 6.

109. *See supra* notes 85–86 and accompanying text.

110. *See supra* Section II.B.1.

111. *See supra* Section II.B.1.

112. *See* BOSTROM, *supra* note 34, at 187. Again, the question of which human interests to protect is a gating item. *See id.* The EP unfortunately did not address this topic, though it does subsequently discuss some specific areas of concern for robotics researchers, including assuring inclusiveness of stakeholders, safety, and privacy. *See EP Report, supra* note 66, at 22–23.

interests of a select few humans, such as their creators, but rather society at large.¹¹³

The critiques levied against the Laws of Robotics apply equally to the EP Principles. Notably, whether these Principles are applied directly to machines or to their researchers, they may still result in similar issues—difficulty interpreting abstract concepts and suboptimality.¹¹⁴ First, these Principles only amplify the abstractness of the Laws of Robotics, now requiring not only that the concept of harm be interpreted, but also human interests, fairness, and autonomy, among others. For instance, humans would disagree regarding what constitutes human interests (wealth, happiness, productivity, recreation, etc.) or fairness (equal benefit to all, equal opportunity, etc.). A machine's researchers could simply pick their own interpretation, and in doing so, input their own biases.¹¹⁵ Second, these Principles, too, would result in significant suboptimality.¹¹⁶ Again, it is unclear whether machines should never harm humans. For instance, beneficence and nonmaleficence may conflict insofar as it is in human interests to eliminate a dangerous human target.¹¹⁷

Note, finally, that the Committee on Legal Affairs also opened the door to imposing significant legal and other ethical obligations onto machines. Particularly, the “guiding ethical framework” must both account for the EP Principles as well as the “principles enshrined in the EU Charter of Fundamental Rights, such as human dignity and human rights, equality, justice and equity, non-discrimination and non-stigmatisation, autonomy and individual responsibility, informed consent, privacy and social responsibility, and on existing ethical practices and codes . . .”¹¹⁸ Thus, the entirety of the constitutional rights of the EU, including such abstract concepts as equality and privacy, may apply to machines and their creators. The framework must also account for case-by-case adjustments to ensure proper behavior in different situations.¹¹⁹

Unlike the Laws of Robotics, then, this approach sacrifices simplicity for comprehensiveness. Rather than solely converging to a few major principles, the Committee on Legal Affairs opts to introduce significantly more legal requirements, ethical practices, and fact-

113. See *EP Report*, *supra* note 66, at 4–5. Conversely, machines could be produced at such scale and low cost that average residents could afford them. With distributed ownership would come distributed benefits. See *id.* at 15.

114. See *supra* Section II.B.1.

115. See *supra* note 55 and accompanying text.

116. See *supra* Section II.B.1.

117. See Marcus, *supra* note 77; *supra* Section II.B.1.

118. *EP Report*, *supra* note 66, at 10.

119. See *id.* at 12.

specific determinations. This kitchen sink approach, however, provides little guidance as to how these various requirements intersect, particularly when or whether legal rights, ethical practices, or case-by-case determinations will trump or conflict with the EP Principles. To that end, the suggestions of the Committee on Legal Affairs may be less of a feasible framework and more a recognition that AI will need to be regulated in responsible ways in the future under EU laws and EP Principles. Rather than untangling the full extent of the kitchen sink, however, this Section has focused on the EP Principles as the more interesting contribution to the realm of AI Laws.

3. Oren Etzioni's AI Rules

In September 2017, Oren Etzioni, chief executive of the Allen Institute for Artificial Intelligence, set out three AI rules with a primarily legal focus.¹²⁰ This Article terms these the “Etzioni Rules.” Although “inspired by” Asimov’s original three Laws of Robotics, Etzioni critiques the Laws of Robotics as “elegant but ambiguous.”¹²¹ His intent is to provide a “more concrete basis for avoiding A.I. harm.”¹²² His “Rules” are as follows:

1. “[A]n A.I. system must be subject to the full gamut of laws that apply to its human operator.”¹²³
2. “[A]n A.I. system must clearly disclose that it is not human.”¹²⁴
3. “[A]n A.I. system cannot retain or disclose confidential information without explicit approval from the source of that information.”¹²⁵

The first Rule is one of depth. A bot must be legally compliant with all laws, just as humans would be, whether operating on behalf of private, corporate, or government actors.¹²⁶ To implement the first Rule, the common law must be amended so that operators are not excused for misconduct solely because the bot acted unexpectedly.¹²⁷ Notably, the first Rule only applies to the laws of human operators, rather than bespoke laws that would apply directly to the bot. Etzioni

120. See Etzioni, *supra* note 47.

121. *Id.*

122. *Id.*

123. *Id.*

124. *Id.*

125. *Id.*

126. *Id.*

127. *Id.*

appears to acknowledge the need to directly regulate bots with the second and third Rules, which do impose unique restrictions on bots.

The second Rule aims to deter the proliferation of falsified content.¹²⁸ Etzioni emphasizes the proliferation of “fake tweets” and “fake news videos.”¹²⁹ For instance, he notes the use of pro-Trump political bots impersonating humans prior to the 2016 US presidential election¹³⁰ that may have influenced the election.¹³¹ Etzioni’s intent appears to be minimizing AI’s ability to fool humans by impersonating humans.¹³² To that end, the second Rule would also generally mitigate bot exploitation of human cognitive biases. AI has already made possible an unprecedented level of personalized interaction with consumers, including the exploitation of those consumers’ vulnerabilities.¹³³ Notably, bots are increasingly assuming human-like characteristics, including human cosmetics, shape, speech patterns, and temperaments.¹³⁴ For instance, Google’s new Duplex AI system (a voicebot) can make restaurant and other reservations by phone, engaging in natural-sounding conversations that are indistinguishable from those with a human.¹³⁵ As apparent “agents,” human-like bots can more easily develop rapport with clients and, in turn, extract more information from clients.¹³⁶

The third Rule intends to prevent bots from misusing user information.¹³⁷ Given bots’ “exceptional ability to automatically elicit, record and analyze information,”¹³⁸ such bots may acquire and misuse confidential information with significantly more ease.¹³⁹ Etzioni’s examples include smart speakers or AI-controlled toys, which could collect private user conversations if permitted by law.¹⁴⁰ Such measures

128. *See id.*

129. *See id.*

130. *See id.*

131. *See* John Markoff, *Automated Pro-Trump Bots Overwhelmed Pro-Clinton Messages, Researchers Say*, N.Y. TIMES (Nov. 17, 2016), <https://www.nytimes.com/2016/11/18/technology/automated-pro-trump-bots-overwhelmed-pro-clinton-messages-researchers-say.html> [<https://perma.cc/42P2-7QNF>].

132. *See* Etzioni & Weld, *supra* note 89, at 18; Etzioni, *supra* note 47.

133. *See* Ryan Calo, *Digital Market Manipulation*, 82 GEO. WASH. L. REV. 995, 999 (2014).

134. *See* Ryan Calo, *Peeping HALs: Making Sense of Artificial Intelligence and Privacy*, 2 EUR. J. LEGAL STUD. 168, 168–69 (2010).

135. *See* Andrew Tarantola, *Pretty Sure Google’s New Talking AI Just Beat the Turing Test*, ENGADGET (May 8, 2018), <https://www.engadget.com/2018/05/08/pretty-sure-googles-new-talking-ai-just-beat-the-turing-test/> [<https://perma.cc/R7UU-BRRG>].

136. *Id.*

137. *See* Etzioni, *supra* note 47.

138. *Id.*

139. *See id.*

140. *See id.*

to protect sensitive information may be even more significant in light of AI's ability to impersonate humans to manipulate consumer behavior.¹⁴¹ Thus, the third Rule dovetails with the second to amplify consumer protections.¹⁴²

The main benefits of the Etzioni Rules are concreteness and comprehensiveness. Moral concepts, particularly of the kind in the Laws of Robotics, tend to be too ambiguous to encode into a machine.¹⁴³ Using specific legal requirements to focus on concrete harms, the Etzioni Rules provide more guidance on the types of prohibited conduct that can be encoded into a machine.¹⁴⁴ Moreover, the Etzioni Rules may prove comprehensive enough to cover largely all bot conduct. By referencing the entire gamut of laws applicable to a human operator, the first Rule in fact applies the entire body of laws to the bot, rather than expecting it to infer how to minimize harm to humans.¹⁴⁵

The Etzioni Rules' comprehensiveness, however, points to the key shortcomings of such a system—added complication and little implementation detail. Unlike the three Laws of Robotics, the Etzioni Rules do not stop at three simple sentences. The first Etzioni Rule makes the entire existing legal system potentially applicable to a sophisticated bot.¹⁴⁶ Bots would need to be able to comply with all domain-specific laws applicable to the activities in which they are engaged. Those laws may also change over time, such that bots must be revisited and upgraded with additional legal knowledge. The first Rule thus has a very different character from the second and third Rules. The former applies a wide array of expectations to each bot that must be determined on a case-by-case basis, whereas the latter imposes two specific legal requirements to which all bots must subscribe.¹⁴⁷ For simplicity, it may be attractive to phrase the Etzioni Rules in the same form as the Laws of Robotics, but they are functionally different by orders of magnitude. The Etzioni Rules unfortunately contain none of the implementation details necessary to create a bot that comprehends so many laws.

Just as importantly, the Etzioni Rules, particularly the first Rule, fail to create a bespoke solution for regulating bots. The first Rule

141. See Calo, *supra* note 133, at 998–99.

142. See Etzioni, *supra* note 47.

143. See *id.*; *supra* Section II.B.1; *infra* Part III.

144. See Etzioni, *supra* note 47. For a discussion of how such laws could be encoded into a machine, see discussion *infra* Part IV.

145. See *infra* Part IV.

146. See *infra* Part IV.

147. See *infra* Part IV.

assumes that bots will operate just like their human operators, such that requiring bots to comply with the laws applicable to their operators will prevent the bot from engaging in nearly all illegal conduct. That may be true in some instances, such as a bot engaging in hacking or processing of sensitive medical information. However, bots come in many forms, from robots to chatbots, voicebots, and sophisticated computer software,¹⁴⁸ and many of the activities in which these bots engage will create unique legal needs. For instance, does it make sense to impose the current civil liability rules on self-driving cars? Many have argued that it would not, proposing such solutions as shifting liability to manufacturers¹⁴⁹ and imposing compulsory insurance schemes.¹⁵⁰ Moreover, bots (even humanoids) do not have human needs. Applying the entirety of human laws to machines would result in nonsensical outcomes, such as machines requiring salaries, overtime pay, or sick days.¹⁵¹ The first Rule, then, can provide some guidance to the types of laws that must be applied to bots, but is not sufficiently tailored to their unique needs.

Ultimately, the Etzioni Rules pose a much-needed reframing of AI Laws in terms of concrete legal rules, reducing the ambiguity of moral AI Laws like Asimov's. However, they focus too little on significant practical questions and fail to address implementation details or the novel legal challenges that bots will present.

4. Google's Technical Guidelines for Intelligent Agents

Rather than focusing on the moral or legal aspects of machine conduct, AI Laws may instead focus on the technical aspects of AI conduct. This approach would start not at abstract thought, but rather at the machine's code, with the goal of minimizing unintended behavior.

Researchers at Google Brain, Open AI, UC Berkeley, and Stanford University have developed such a set of technical rules.¹⁵²

148. See *supra* notes 2–13 and accompanying text.

149. See, e.g., Kyle Colonna, *Autonomous Cars and Tort Liability*, 4 CASE W. RES. J. L. TECH. & INTERNET 81, 83 (2012); Gary E. Marchant & Rachel A. Lindor, *The Coming Collision Between Autonomous Vehicles and the Liability System*, 52 SANTA CLARA L. REV. 1321, 1326 (2012); Andrew P. Garza, Note, "Look Ma, No Hands!": *Wrinkles and Wrecks in the Age of Autonomous Vehicles*, 46 NEW ENG. L. REV. 581, 583 (2012).

150. See *EP Report*, *supra* note 66, at 60–61; Taylor Wessing, *Who's in the Driving Seat? Driverless Cars, Liability and Insurance*, LEXOLOGY (Nov. 3, 2017), <https://www.lexology.com/library/detail.aspx?g=40b3a3ea-4e8d-4e3b-8c82-a7a5e3b44aad> [<https://perma.cc/5M2K-R97K>].

151. See, e.g., Federal Labor Standards Act, 29 U.S.C. §§ 206–207 (2018) (requiring federal minimum wage and overtime pay); CAL. LABOR CODE § 246 (West 2018) (requiring sick leave).

152. See Amodעי et al., *supra* note 48, at 21.

Although the media widely hailed this research as Google's attempt to create laws of robotics,¹⁵³ it is indeed quite the opposite. Rather than programming bots to act in a morally or legally desirable way, the researchers focused on addressing "accidents in machine learning systems,"¹⁵⁴ or unintended behaviors associated with "machine learning-related implementation errors."¹⁵⁵ Their research also eschews extreme, speculative, and doomsday scenarios of AI misconduct and focuses instead on practical errors in machine learning that can be applied and experimented with today.¹⁵⁶ In this sense, Google's research may be thought of more as a set of procedures or best practices for programmers, rather than a particular set of substantive values or norms. Finally, their research is primarily concerned with reinforcement learning,¹⁵⁷ a specific paradigm of machine learning in which machines learn on their own with no prior data by repeatedly performing tasks.¹⁵⁸ By honing in on one paradigm, the researchers were able to tailor solutions to particular sets of problems.

Specifically, the researchers developed the following set of guidelines (referred to as the "Google Guidelines" in this Article) to address five basic problems for building responsible bots engaged in reinforcement learning:

1. Avoiding Negative Side Effects: How can a machine be prevented from affecting its environment in negative ways while performing the tasks it is given?¹⁵⁹
2. Avoiding Reward Hacking: Given that a machine is built to maximize a reward function, how can it be prevented from gaming that function?¹⁶⁰

153. See, e.g., John Brownlee, *Google Created Its Own Laws of Robotics*, FAST COMPANY (June 24, 2016), <https://www.fastcodesign.com/3061230/google-created-its-own-laws-of-robotics> [<https://perma.cc/F8WZ-6BUD>]; Ben Lovejoy, *Google Formulates Real-Life Version of Asimov's Three Laws of Robotics for Safe & Reliable AI*, 9TO5GOOGLE (June 22, 2016), <https://9to5google.com/2016/06/22/google-safe-ai-rules/> [<https://perma.cc/ZZ4Q-KNRX>].

154. See AMODEI ET AL, *supra* note 48, at 1.

155. *Id.* at 1–2.

156. *See id.* at 2.

157. *See id.* at 3–4.

158. See Will Knight, *Reinforcement Learning*, MIT TECH. REV., <https://www.technologyreview.com/s/603501/10-breakthrough-technologies-2017-reinforcement-learning/> [<https://perma.cc/X9HQ-KVNM>] (last visited Sept. 16, 2018). By experimenting, the machine develops the large dataset it needs to then optimize its behavior. *See id.*

159. See Amodei et al., *supra* note 48, at 2.

160. *See id.*

3. Scalable Oversight: How can a machine accomplish its tasks properly even if it has only limited access to human input and evaluation?¹⁶¹
4. Safe Exploration: How can the machine experiment to improve its performance while limiting negative consequences of such experimentation?¹⁶²
5. Robustness to Distributional Shift: How can a machine recognize different real-world environments and behave appropriately in each environment?¹⁶³

The researchers then proposed potential solutions in each such area.¹⁶⁴ They focus on testability, presenting a potential experiment for every problem area.¹⁶⁵ They conclude that AI researchers need a “unified approach to prevent [autonomous] systems from causing unintended harm” in lieu of the current system of “ad hoc fixes.”¹⁶⁶

Some may justifiably argue that the Google Guidelines do not constitute a definitive set of comprehensive AI Laws. The research does not provide one consistent set of rules for all intelligent machines. Furthermore, the researchers specifically do not address all types of bots,¹⁶⁷ nor do they propose a definitive way to minimize problems in each of the areas they explore.¹⁶⁸ The Google Guidelines thus lack the simplicity of Asimov’s Laws of Robotics or Etzioni’s AI Rules, both of which define a clear set of comprehensive mandates.

That, however, is exactly the charm of the Google Guidelines. Rather than discuss these problems at a philosophical level, the researchers engaged in an empirical study,¹⁶⁹ resulting in a framework that could be used to identify and address a “broad variety of potential risks, both short- and long-term.”¹⁷⁰ Moreover, although no definitive answer was provided, the researchers did provide more certainty than past philosophical proposals—insofar as they discussed the

161. See *id.* at 3.

162. See *id.*

163. See *id.*

164. See *id.* at 4–20. For instance, to mitigate reward hacking, the researchers proposed implementing trip wires, capping rewards, offering multiple rewards, and careful engineering. *Id.* at 10–11. However, the recommendations for each of the five problems are too voluminous and technical to discuss in this Article.

165. See *id.* at 7, 11, 13, 15–16, 20.

166. *Id.* at 21.

167. See *id.*

168. See *id.* (“We presented five possible research problems related to accident risk and for each we discussed possible approaches that are highly amenable to concrete experimental work.”).

169. See *id.* at 20.

170. *Id.*

implementation details of their solutions.¹⁷¹ The Google Guidelines address known problems that researchers already face and present solutions that can be programmed and tested today.¹⁷² In contrast, the Laws of Robotics are currently beyond verification because we do not have sufficiently advanced machines that could comply with general principles—and we may never.¹⁷³

The solutions proposed in this Article follow this technically-driven set of Guidelines. Only by understanding and designing behaviors around the technological limits of AI can robust systems be created to handle a wide variety of risks. Implementation details must be at the forefront of any real-world solution.

C. Conclusion: Revisiting the Proposals of AI Laws

This Article has examined four sets of AI Laws: (1) Asimov's Laws of Robotics, (2) the EP Principles, (3) the Etzioni Rules, and (4) the Google Guidelines. Most significantly, this Part has demonstrated the multitude and variance among proposals, ranging from highly abstract moral tenets (Asimov's Law of Robotics and the EP Principles) to concrete legal rules (the Etzioni Rules) to technical guidelines for minimizing unexpected bot conduct (the Google Guidelines). Table 1 summarizes these AI Laws:

171. See *id.* at 7, 11, 13, 15–16.

172. See *id.* at 2.

173. See BOSTROM, *supra* note 34, at 19 (indicating a possibility that an AGI will never be developed); Marcus, *supra* note 77.

Table 1. AI Laws Summary

AI Laws	Domain	Pros	Cons
Laws of Robotics	Moral	Simple	Too abstract; simultaneously over-inclusive and under-inclusive; internally contradictory
EP Principles	Moral	Comprehensive	Too abstract; simultaneously overinclusive and underinclusive; complex (scope poorly defined)
Etzioni Rules	Legal	Comprehensive; concrete	No implementation detail; does not account for bot uniqueness and code
Google Guidelines	Technical	Comprehensive; testable; accounts for bot uniqueness and code	Open-ended

III. MORALITY AS A POTENTIAL DOMAIN OF AI LAWS

Part II demonstrates the various ways in which AI Laws have been conceptualized—with some emphasizing moral, legal, or technical factors. However, the most significant proposal—the Laws of Robotics—and similarly inspired laws, operate at the level of morality. Prior to developing a more effective set of AI Laws, it bears asking: is morality the proper domain for AI Laws?

This Part analyzes the benefits and downsides of using morality as a framework to control bots and concludes that morality could not generate a feasible, enforceable set of AI Laws. Rather, AI Laws must be grounded in more specificity. As Part IV discusses, that specificity can be found in a system of laws.

A. Benefits of Moral AI Laws

Given the large number of bots predicted to be deployed in the future,¹⁷⁴ it may be impossible to monitor all their actions. Rather than codify the many sets of laws across every jurisdiction and area of law,¹⁷⁵ attention can be given to a few well-defined tenets—particularly for moral sets of laws in the vein of Asimov’s. By creating one common, simple ethical system, society can at least ensure that it builds in the proper moral sense to allow bots to adapt to the various situations that they will face. To that end, some have argued that programming a bot to mindlessly comply with every law could increase potential dangers,¹⁷⁶ leaving it unable to adapt to its then-current environment. For instance, a self-driving car may need to exceed the speed limit where an emergency so requires.¹⁷⁷ Moreover, bots would be monitored and programmed for compliance with only one set of principles, rather than various legal codes, which would potentially ease the administrative burden of controlling each bot.

Developers can arguably program precise decisions into their bots to allow those bots to account for and minimize liability under current laws.¹⁷⁸ However, developers also face more responsibility for their decisions.¹⁷⁹ Unlike humans, who must make real-time decisions on little notice, bots are programmed in advance, and so developers have the time to determine appropriate behaviors for their bots.¹⁸⁰ Rather than relying on developers to conceive of and develop technical solutions to every potential legal issue, a transparent system of bot ethics may more adequately account for the sophistication of bots and generate better outcomes for society.

174. See *supra* note 14 and accompanying text.

175. See *infra* Section V.A.

176. See, e.g., Patrick Lin, *The Ethics of Autonomous Cars*, ATLANTIC (Oct. 8, 2013), <https://www.theatlantic.com/technology/archive/2013/10/the-ethics-of-autonomous-cars/280360> [<https://perma.cc/Y2HZ-BKRR>] (“We need to ensure [self-driving cars] make moral sense. Programming a robot car to slavishly follow the law, for instance, might be foolish and dangerous.”). Note, however, that Lin does not account for the doctrine of necessity, by which laws may be breached where a greater social good so requires. For a discussion of the definition and elements of necessity, see Edward B. Arnolds & Norman F. Garland, *The Defense of Necessity in Criminal Law: The Right to Choose the Lesser Evil*, 65 J. CRIM. L. & CRIMINOLOGY 289, 289–95 (1974).

177. Lin, *supra* note 176.

178. See *id.*

179. See *id.*

180. See *id.*

B. Downsides of Moral AI Laws

However, such potential benefits likely cannot be realized because encoding morality into a machine is both technically and practically difficult—if not impossible—to implement. Moral principles typically tend to be ambiguous, and thus difficult to “translate into precise system and algorithm design.”¹⁸¹ Just as there may be no one definitive concept of “harm,”¹⁸² there may also be no way to encode such a concept, given its multitude of meanings. To choose one meaning would prioritize one interpretation over others—encoding only one aspect of such a concept.

The ethics approach is also fundamentally flawed insofar as there is no one optimal system of ethics.¹⁸³ Many ethical systems have been proposed for intelligent machines, including machine ethics, computer ethics, robot ethics, and machine morals.¹⁸⁴ Most ethical arguments in favor of such systems have, however, tended to be philosophical, largely arguing about which set of ethics—utilitarianism, Kantianism, etc.—should be mechanized.¹⁸⁵ Agreeing on one ethical code is impossible because there is no universally approved system of ethics¹⁸⁶ as ethics “vary according to culture, religion, and beliefs.”¹⁸⁷ Even a local communal standard would be difficult to administer, particularly regarding the specifics of each moral concept.¹⁸⁸ For instance, people may agree about what harms to prohibit in the abstract (such as protecting the environment), but not the specific implementation details (such as fuel emissions standards in vehicles).¹⁸⁹ These machines will face moral dilemmas as they attempt to manage contradictory values.¹⁹⁰

Furthermore, imbuing a machine with ethics would largely imbue it with one’s own values and biases. AI’s susceptibility to bias has been widely criticized,¹⁹¹ as bots “replicate and magnify” the biases

181. EXEC. OFFICE OF THE PRESIDENT, NAT’L SCI. & TECH. COUNCIL, PREPARING FOR THE FUTURE OF ARTIFICIAL INTELLIGENCE 28 (2016).

182. See *supra* Section II.B.

183. See Yampolskiy, *supra* note 45, at 389.

184. See *id.*

185. See *id.*

186. *Id.*

187. NSTC PLAN, *supra* note 55, at 27.

188. Amitai Etzioni & Oren Etzioni, *AI Assisted Ethics*, 18 ETHICS & INFO. TECH. 149, 151 (2016).

189. *Id.*

190. See NSTC PLAN, *supra* note 55, at 27.

191. See, e.g., CATHY O’NEIL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY 199–200 (2016); Crawford, *supra* note 57.

of their creators.¹⁹² Such biases may be introduced into a bot through one's choice of dataset or algorithm. For instance, a bot that processes internet-based text may reflect gender stereotypes,¹⁹³ or crime prediction software may assign higher risk to racial minorities.¹⁹⁴ Similarly, a bot that adopts one interpretation of a nuanced ethical concept would reflect its creator's interpretation. Such biases would, moreover, be difficult to detect and correct. Since humans trust the apparent objectivity of machines, they tend to be less likely to notice the human biases input into machines.¹⁹⁵

The ethics approach also creates significant resource constraints. The law is normative insofar as it reflects a specific community's values.¹⁹⁶ However, the types of values which are enshrined in the law do have limits.¹⁹⁷ Certain individuals may disagree with the widespread sale of alcohol and tobacco, employers who yell at their employees, or bad customer service—yet we do not legally punish individuals for any of those, and many people would see no issue with any of those activities. Punishing such ubiquitous activities would create problems of enforceability by requiring significant increases to law enforcement and legal resources, such as labor, prison space, and legal expertise.¹⁹⁸ Monitoring bots for morality will likely result in similar burdens.

C. Relative Morality: A Different Form of Moral AI Laws

One of the primary critiques levied against morality is the lack of a single, optimal moral code.¹⁹⁹ It then bears asking whether individualized moral codes could remedy this deficiency. That is, how does a system of relative morality compare to a system of universal morality?

192. See AI NOW 2016 REPORT, *supra* note 56, at 6; EXEC. OFFICE OF THE PRESIDENT, *supra* note 181, at 30 (“AI needs good data. If the data is incomplete or biased, AI can exacerbate problems of bias.”); Crawford, *supra* note 57.

193. See CAMPOLO ET AL., *supra* note 57, at 18–19.

194. See Angwin et al., *supra* note 57.

195. See CAMPOLO ET AL., *supra* note 57, at 30.

196. See LAWRENCE M. FRIEDMAN, LAW IN AMERICA: A SHORT HISTORY 17 (2004).

197. See *id.*

198. See Richard A. Posner, *The Costs of Enforcing Legal Rights*, 4 E. EUR. CONST. REV. 71, 72 (1995) (identifying potential resource costs for enforcing liberties, including police, prosecutors, judges, and subsidized lawyers).

199. See Yampolskiy, *supra* note 45, at 389. For several predictions relating to the extensive scope of AI automation from reputable firms, see Press, *supra* note 74.

Universal morality would ascribe an overarching ethical system to all bots, such that all have the same moral precepts.²⁰⁰ All bots would be provided with the same moral code so that all behave similarly in moral dilemmas. Asimov's Laws of Robotics provide an example of such a system. The underlying premise for a universal morality system is that one consistent ethical system exists to generate socially desirable behavior.²⁰¹ The discussions of moral AI Laws throughout this Article have focused primarily on universal morality.

Conversely, relative morality would grant bots the same moral precepts as their operators so that the bots act consistently with their operators' preferences.²⁰² All bots would be provided with the same moral code as their owners.²⁰³ Acknowledging that bots should be programmed with both a sense of law and values, Amitai and Oren Etzioni propose that a bot's moral precepts should come directly from its operator, rather than society at large.²⁰⁴ In this perspective, the law enshrines communal values, which all citizens must uphold—whereas ethics govern personal values, which are properly left to the discretion of each individual.²⁰⁵

Implementation details pose a special difficulty for relative morality, particularly with respect to how a bot should decipher and understand an individual's moral preferences. Amitai and Oren Etzioni also propose use of AI oversight bots, or "AI Guardians," that oversee operations of other active bots and guarantee that they comply with their design guidelines.²⁰⁶ Notably, they argue in favor of implementing one type of guardian—an ethics bot—that would be tasked with deciphering a person's ethical preferences by processing both a person's public information on the internet and data within a personal computer.²⁰⁷ Those preferences would then be fed into other bots to allow those bots to operate in accordance with their operator's moral

200. See Gilbert Harman, *Moral Relativism Defended*, 84 PHIL. REV. 3, 3 (1975); *supra* Section II.B.1.

201. *Moral Universalism*, BASICS OF PHIL., https://www.philosophybasics.com/branch_moral_universalism.html [https://perma.cc/38ZB-GP9B] (last visited Sept. 18, 2018).

202. See Harman, *supra* note 200, at 3.

203. See Amitai Etzioni & Oren Etzioni, *Designing AI Systems That Obey Our Laws and Values*, 59 COMM. OF THE ACM 29, 31 (2016).

204. *Id.*

205. See *id.*; Etzioni & Etzioni, *supra* note 188, at 151.

206. Etzioni & Etzioni, *supra* note 203, at 29–30.

207. *Id.* at 31.

preferences.²⁰⁸ AI Guardians would also guarantee compliance with legal and ethical norms.²⁰⁹

Relative morality largely reflects how the world already operates. Individuals may hold and express their moral preferences however they wish. They must make moral choices whenever managing employees, treating clients, or making consumption decisions that affect the environment. As bots act as agents in traditionally human activities, those bots would act like the humans they replace—that is, mimic decisions of their owners or operators.²¹⁰

Notably, relative morality would also amplify the very risks that bots seek to mitigate. Many have emphasized AI's transformative role in ensuring social safety, such as by reducing up to 90 percent of traffic accidents.²¹¹ Nevertheless, by adopting unique moral codes, nearly all bots will learn to behave differently, according to the preferences of their operators. That, in turn, will decrease predictability and capacity for oversight, with bots working to accomplish different goals.

Moreover, a system of relative morality would likely prove even less implementable than a system of universal morality. Most notably, such a system would result in significantly more inconsistency and ambiguity, as ethics would need to be determined on a case-by-case basis for the billions of then-living individuals. Bots already struggle with concepts like “harm,”²¹² but relative morality would amplify that ambiguity exponentially—attempting to tackle the innumerable ethical concepts and idiosyncratic definitions that the global population ascribes to them. Individuals may, moreover, hold inconsistent ethical values or practice these values inconsistently.²¹³ An ethics bot would effectively decide how to address such situations, rather than reflecting how its conflicted operator would handle the situation in the moment.²¹⁴ Over a longer span of time, individuals' preferences may also gradually change. A bot, then, must both understand abstract, vague concepts and continually update those concepts to properly reflect its operator's intention.

208. *Id.*

209. *See id.* at 30–31.

210. *See id.*

211. Adrienne Lafrance, *Self-Driving Cars Could Save 300,000 Lives Per Decade in America*, ATLANTIC (Sept. 29, 2015), <https://www.theatlantic.com/technology/archive/2015/09/self-driving-cars-could-save-300000-lives-per-decade-in-america/407956/> [https://perma.cc/Q74C-RPSK].

212. *See* Marcus, *supra* note 77.

213. *Consistency and Ethics*, SANTA CLARA U. (Jan. 1, 1998), <https://www.scu.edu/ethics/ethics-resources/ethical-decision-making/consistency-and-ethics/> [https://perma.cc/FZH3-LTCR].

214. Etzioni & Etzioni, *supra* note 188, at 153–54.

Relative morality may also be technically infeasible, particularly with respect to its use of ethics bots. Amitai and Oren Etzioni mention that such an ethics bot would “analyze[] many thousands of items of information.”²¹⁵ However, they may be overestimating the sheer quantity of moral data that personal computers and publicly available information may present. People do not generally take to Facebook and exclaim, “If an armed intruder trespassed into my home, I would use lethal force to defend myself!” Similarly, a person’s computer applications, privacy settings, and browser history may not reveal much about that person’s views on customer service, use of prescription medications, or lying to avoid hurting a friend’s feelings. Data of sufficient granularity may not exist.

Data quantity and quality issues would thus abound, so much so that sufficiently good results likely could not be generated. Modern bots use millions to trillions of data points or more, rather than thousands.²¹⁶ Without such large quantities of data, bots in fact tend to perform rather poorly.²¹⁷ Moreover, a bot must be trained with high-quality data that accurately represents the many potential options it is attempting to learn or predict.²¹⁸ Acquiring sufficient representative data would present special difficulty for areas such as an individual’s ethical views, where potentially countless options exist. In the absence of such high quantity and quality of data, an ethics bot would fail to grasp the full spectrum of its human operator’s moral preferences and feed incorrect preferences into other bots. As a result, the actions of bots and their human operators would deviate.

D. Conclusion

This Part demonstrates the inadequacy of morality as the realm for AI Laws, both on technical and practical grounds. On the technical front, such bots would be imbued with ambiguous, deficient moral principles that could generate contradictory results. On the practical front, implementing such morals would prove difficult, given the countless potential ethical systems that could be encoded. Encoding any particular system of ethics would not only favor one particular

215. *Id.* at 152.

216. *See* Kaasschieter, *supra* note 25; Metz, *supra* note 5.

217. Nello Cristianini, *The Road to Artificial Intelligence: A Case of Data over Theory*, *NEW SCIENTIST* (Oct. 26, 2016), <https://www.newscientist.com/article/mg23230971-200-the-irresistible-rise-of-artificial-intelligence/> [<https://perma.cc/LLB9-S6AQ>]; Bernard Marr, *Why AI Would Be Nothing Without Big Data*, *FORBES* (June 9, 2017, 12:29 AM), <https://www.forbes.com/sites/bernardmarr/2017/06/09/why-ai-would-be-nothing-without-big-data/#1cfc793e4f6d> [<https://perma.cc/W7FT-G8UQ>].

218. *See* Kaasschieter, *supra* note 25.

worldview, but potentially waste resources that could be committed to making safe, legally compliant bots. As Part IV discusses, the solution to the technical problem lies in implementing a clearer, more certain set of machine-interpretable laws. As Part V discusses, the solution to the practical problem lies in developing compartmentalized modules of code to track particular sets of laws, which can then be easily integrated into bots.

IV. A COMPUTING MODEL OF LEGAL ABSTRACTION: MAKING LAWS MACHINE INTERPRETABLE

With morality being an unattractive domain for AI Laws, the question becomes whether the law would fare any better. How can computers think like lawyers?

As Cass Sunstein has argued, bots may not even have the capacity to engage in legal reasoning, which often involves analogical reasoning that has an “evaluative, value-driven character.”²¹⁹ Lawyers must both understand the facts of a case or transaction (the specifics) and then abstract from those facts (generalize). As discussed, computers currently fail to engage in high-level abstraction.²²⁰

However, the law operates at a lower level of abstraction than morality. The law contains rules and standards. The former tend to be clear-cut and unambiguous.²²¹ The latter tend to be ambiguous and ad hoc but, even so, are not decided in a vacuum.²²² Rather, the facts and circumstances of a case must provide enough fodder for judges to decide the outcome of a standard.²²³ Thus, the law manages abstraction both on a practical and a technical level insofar as legal concepts tend to be less ambiguous than highly abstract thought, and cases occur in the context of (often) significant amounts of facts and case law, respectively.

219. Cass R. Sunstein, *Of Artificial Intelligence and Legal Reasoning* 5 (U. Chi. Pub. L. & Legal Theory, Working Paper No. 18, 2001), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=289789 [<https://perma.cc/E9KG-7QBW>]. Note, however, that Sunstein specifically claims that AI may reach such capacity in the future. *Id.* at 8. Seventeen years after his paper was published, there are several legal-tech companies trying to build a robot lawyer that can, at the very least, find legal solutions by identifying patterns in case law. See, e.g., Steve Lohr, *A.I. Is Doing Legal Work. But It Won't Replace Lawyers, Yet.*, N.Y. TIMES (Mar. 19, 2017), <https://www.nytimes.com/2017/03/19/technology/lawyers-artificial-intelligence.html> [<https://perma.cc/D5XL-WPX9>].

220. See Marcus, *supra* note 77.

221. Russell B. Korobkin, *Behavioral Analysis and Legal Form: Rules vs. Standards Revisited*, 79 OR. L. REV. 23, 25 (2000).

222. *Id.* at 25–26.

223. *Id.*

This Part demonstrates bots' capacity to engage in legal reasoning in three sections. First, it addresses how bots operate, specifically focusing on two factors—computer code and AI learning. Then, it addresses how laws operate, namely by analyzing clear-cut legal rules, ambiguous legal standards, and abstract extralegal matters like ethics. Finally, it draws a parallel between the two, demonstrating that legal rules can be encoded through direct instruction and standards learned through data analysis.

A. *How Bots Think: A Computer's Instruction and Data Components*

A bot can be directed to perform a task in two ways—explicit instruction or data.²²⁴ For simplicity, this Article refers to each of these as a “component.” Explicit instruction refers to traditional programming, whereby humans write code into the computer that specifies exactly what the computer should do. Conversely, a bot can also learn to act in more nuanced, abstract ways by processing large datasets that teach it how to process new, but similar, data. For instance, a bot may learn to speak like a human by processing voice data or drive by processing driving data, even though coding it to speak or drive like a human would be infeasible.²²⁵ Note that these two components are intimately intertwined in any bot: the bot requires explicit coding to direct it to learn from data, and it ultimately reduces data to explicit instructions that it can understand and execute.²²⁶ This Section examines each of these components in more depth.

1. Explicit Commands Through the Instruction Component

Current computers must be directed to act through explicit commands, which are issued to the computer through machine code.

224. See Jeff Leek, *What Is Artificial Intelligence? A Three Part Definition*, SIMPLY STAT. (Jan. 19, 2017), <https://simplystatistics.org/2017/01/19/what-is-artificial-intelligence/> [<https://perma.cc/X7PK-JT8P>]. Leek defines AI as comprising three components: “1. *The data set*: A [set] of data examples that can be used to train a statistical or machine learning model to make predictions[.] 2. *The algorithm*: An algorithm that can be trained based on the data examples to take a new example and execute a human-like task[.] 3. *The interface*: An interface for the trained algorithm to receive a data input and execute the human like task in the real world.” *Id.* The discussion in this Article focuses on the first two of these components.

225. See *id.* Theoretically, these types of difficult tasks could be directly coded into a computer, but would require such lengthy, detailed code and resources that doing so would be nearly impossible. When a computer learns how to perform a task through data analysis, it effectively transforms its nuanced processing of the data into machine code that it can execute. Whether a computer is explicitly instructed or learns through data, it will still translate its tasks into a set of machine code. See *Machine Code Definition*, LINUX INFO. PROJECT (May 11, 2006), http://www.linfo.org/machine_code.html [<https://perma.cc/QN3M-HGJ5>].

226. See Leek, *supra* note 224 (noting that data trains the algorithm).

Machine code, or machine language, refers to the code that a computer's central processing unit executes.²²⁷ It is the language underlying all computer activities, as all programs must be converted into machine code for the computer to execute them.²²⁸ Machine code is difficult for humans to comprehend, use, and debug.²²⁹ To mitigate such issues, developers generally code with simpler programming languages that combine, redefine, and make readable various machine instructions.²³⁰ The more these languages abstract away from machine code, the more "high-level" they are.²³¹

The closer to machine code, the more "low-level" they are.²³² Similarly, developers can also build frameworks, tools, or even entirely new programming languages on top of other languages, which yet again serves to abstract and create higher-level software.²³³

Programming languages and tools thus already allow for some degree of abstraction, by simplifying, combining, and streamlining what the computer can already do. In this sense, explicit coding does not teach an old computer "new tricks," like analogical reasoning.²³⁴

227. See *Machine Code Definition*, supra note 225. For a more extensive discussion of the various types of languages and their levels of abstraction, see Brijender Kahanwal, *Abstraction Level Taxonomy of Programming Language Frameworks*, 3 INT'L J. PROGRAMMING LANGUAGES & APPLICATIONS 1, 2 (2013).

228. *Machine Code Definition*, supra note 225.

229. Kahanwal, supra note 227, at 3.

230. See *id.* at 6–8 (discussing middle-level and high-level programming languages).

231. See *id.* Examples include nearly any commonly used programming language, like Python, C++, or JavaScript. See *Machine Code Definition*, supra note 225. Note, however, that even amongst these languages, some will be considered higher-level than others. Python, for instance, is higher-level than C++ because Python abstracts away more details of machine code. See Kahanwal, supra note 227, at 6–11; *Machine Code Definition*, supra note 225.

232. See Kahanwal, supra note 227, at 3.

233. See *id.* A framework is a module (or set of modules) built on top of a programming language to simplify tasks. Arvind Rongala, *Applications of Python in the Real World*, INVENSIS (Mar. 15, 2015), <https://www.invensis.net/blog/it/applications-of-python-in-real-world/> [<https://perma.cc/F746-DWEB>]. Leading AI frameworks, for example, are built on a variety of programming languages and mitigate the need to develop bespoke neural networks or other machine learning algorithms. See, e.g., Maruti Techlabs, *8 Best Deep Learning Frameworks for Data Science Enthusiasts*, MEDIUM (Apr. 5, 2018), <https://medium.com/the-mission/8-best-deep-learning-frameworks-for-data-science-enthusiasts-d72714157761> [<https://perma.cc/Q3CU-JCTP>].

234. See MARIUSZ FLASIŃSKI, INTRODUCTION TO ARTIFICIAL INTELLIGENCE 16 (2016). For purposes of this discussion, I focus on neural networks and other bottom-up approaches to AI, which are based on learning and generalizing from data. Neural networks, however, come from one school of thought, connectionism. See *Connectionism*, STAN. ENCYC. OF PHIL. (Feb. 19, 2015), <https://plato.stanford.edu/entries/connectionism/> [<https://perma.cc/T7DV-GVRN>]. For further information on connectionism, see *id.* There are other schools of thought. See, e.g., PEDRO DOMINGOS, THE MASTER ALGORITHM: HOW THE QUEST FOR THE ULTIMATE LEARNING MACHINE WILL REMAKE OUR WORLD 51 (2015). Pedro Domingos, for instance, identifies five tribes of machine learning: symbolists, connectionists, Bayesians, analogizers, and evolutionaries. *Id.* Notably, the symbolists represent a top-down approach, trying to engineer computers with the capability to understand highly abstract thought and thus deduce information from first

Bots can be directed to act, just like any computer program. A bot is simply a computer program that is directed to process certain types of data, learn, and perform related tasks. Similarly, if the program's code prohibits it from engaging in certain conduct (like driving over sixty-five miles per hour or recording a user's conversations with the bot), the bot must follow that code.²³⁵ However, human error—whether bad coding or deliberate sabotage—can still result in bots engaging in unintended or unlawful activities.²³⁶

2. Learning Through the Data Component

A machine cannot abstract the same way that a human can. Even modern AI applications learn by processing millions to trillions of data points or more, deciphering intricate patterns along the way.²³⁷ Some have claimed that such processing and pattern recognition make these AIs comparable to humans.²³⁸ However, human and AI problem-solving and comprehension operate quite differently.

principles. *See id.* at 52. If such symbolic manipulation were possible, then a computer program could indeed teach a computer new tricks, including the innate ability to abstract without data. *See* Lea Winerman, *Making a Thinking Machine*, AM. PSYCHOL. ASS'N (Apr. 2018), <http://www.apa.org/monitor/2018/04/cover-thinking-machine.aspx> [<https://perma.cc/WAF8-PUTX>].

235. Babcock et al., *supra* note 67, at 2. This ignores the possibility that a superintelligent program could learn to override its own code. It is currently unclear whether such a self-determining bot is possible. Such a bot would (at a minimum) require significantly more controls to ensure that it acts as intended, such as containment measures. *See id.* However, a self-determining bot may also simply be uncontrollable such that no ex ante restrictions on its capabilities—including instructions to comply with any law—would be feasible. Vernor Vinge, *The Coming Technological Singularity: How to Survive in the Post-Human Era*, NASA CONF. PUBLICATION 10129 11, 12 (Mar. 1993).

236. *See, e.g.*, Hiroko Tabuchi & Jaclyn Trop, *Toyota Recalls Newest Priuses over Software*, N.Y. TIMES (Feb. 12, 2014), <https://www.nytimes.com/2014/02/13/business/international/toyota-issues-another-recall-for-hybrids-this-time-over-software-glitch.html> [<https://perma.cc/8939-LAYX>].

237. *See* Kaasschieter, *supra* note 25.

238. *See, e.g.*, Larry Hardesty, *Explained: Neural Networks*, MIT NEWS (Apr. 14, 2017), <http://news.mit.edu/2017/explained-neural-networks-deep-learning-0414> [<https://perma.cc/UE8L-KWST>]. AI's "humanness" is often measured through the "Turing Test," or the "Imitation Game," proposed by Alan Turing in 1950. *See* A.M. Turing, *Computing Machinery and Intelligence*, 59 MIND 433, 433 (1950). Turing proposed that an AI exhibiting human-like sentience, measured by its ability to fool humans into believing that it is also human, would indeed be deemed sentient. *See id.* at 434; *Turing Test*, ENCYC. BRITANNICA, <https://www.britannica.com/technology/Turing-test> [<https://perma.cc/XQB3-S3JD>] (last visited Sept. 18, 2018). Some AIs may have already passed the Turing Test. *See, e.g.*, *Computer Simulating 13-Year Old Boy Becomes First to Pass Turing Test*, GUARDIAN (June 9, 2014), <https://www.theguardian.com/technology/2014/jun/08/super-computer-simulates-13-year-old-boy-passes-turing-test> [<https://perma.cc/S6GV-NZ4Z>]; Tarantola, *supra* note 135.

Humans intuitively grasp high-level, abstract thoughts.²³⁹ For instance, when solving new problems, we do not need to process or iterate through millions of data points or test every potential outcome—rather, we hypothesize that certain solutions will work well and test those solutions.²⁴⁰ We even succeed at “one-shot learning”—learning to identify concepts like letters of the alphabet after seeing only one example.²⁴¹ That is not to minimize the years of experience, education, and social learning that goes into developing the human mind. However, the human mind seems particularly useful for abstraction, even creating new concepts or solutions.

Bots, in contrast, begin at the lowest level of instruction and work towards high-level thoughts.²⁴² A neural network—the current leading AI paradigm²⁴³—processes data through nodes, or information processing units, that are organized in layers.²⁴⁴ Each network may have anywhere from a few to hundreds of layers, with each becoming increasingly sophisticated.²⁴⁵ For instance, assume a simple neural network detects whether an image contains a human.²⁴⁶ The initial inputs (a dataset that contains images labeled as human or nonhuman) might be passed into a layer that simply detects horizontal, vertical, and diagonal edges in each image. Each node in that layer may, in turn, be responsible for one type of edge: horizontal, vertical, or diagonal. Subsequent layers would then detect geometric shapes, and still later layers would detect features of the human like hands and ears. The final layer would combine all this processing to determine whether a human exists in the image. This granularity is why, for example,

239. See Alex Chavers, *Artificial Intelligence Just Can't Top Human Intuition. Yet.*, NEWS LEDGE (Jan. 9, 2017), <https://www.newsledge.com/artificial-intelligence-cant-top-human-intuition/> [<https://perma.cc/Y2ZZ-MDJS>]; Gary Marcus, *Artificial Intelligence Is Stuck. Here's How to Move It Forward.*, N.Y. TIMES (July 29, 2017), <https://www.nytimes.com/2017/07/29/opinion/sunday/artificial-intelligence-is-stuck-heres-how-to-move-it-forward.html> [<https://perma.cc/4Q9B-AN8C>].

240. See Chavers, *supra* note 239.

241. Winerman, *supra* note 234; see also Brenden M. Lake et al., *Human-Level Concept Learning Through Probabilistic Program Induction*, 350 SCI. 1332, 1322 (2015).

242. See Lake et al., *supra* note 241, at 1332.

243. Brenden M. Lake et al., *Building Machines That Learn and Think Like People*, 40 BEHAV. & BRAIN SCI. 1, 1–2 (2017).

244. Hardesty, *supra* note 238.

245. Lake et al., *supra* note 243, at 22.

246. For a visual example explaining how a neural network performs image recognition, see Shafeen Tejani, *Machines That Can See: Convolutional Neural Networks*, FROM BITS TO BRAINS (Dec. 20, 2016), <https://shafeentejani.github.io/2016-12-20/convolutional-neural-nets/> [<https://perma.cc/K4B5-DCB4>].

machines are better at identifying species of dogs than humans are²⁴⁷—machines can process the details in the many species of dogs that may not be noticeable to the human eye. Yet, machines still have trouble recognizing simple arrangements of geometric shapes.²⁴⁸ This granularity is also why machines do not—and may never have the capacity to—understand highly abstract thoughts like “love.”²⁴⁹ How would a programmer break down “love” into hundreds of components?

That is not to say that bots can only grasp concrete concepts, but rather that their ability to comprehend abstractness is limited. A machine will generalize from data if and only if large datasets exist and those datasets indicate recognizable patterns.²⁵⁰ Stated differently, well-performing AI requires two factors: (1) many data points and (2) identifiable patterns in those data points.²⁵¹ The first is a practical issue. Insofar as data exist but sufficient amounts cannot be collected due to resource constraints, learning is possible but the bot simply does not have the resources to do so.²⁵² The second is a technical issue. If a

247. Ophir Tanz, *Can Artificial Intelligence Identify Pictures Better than Humans?*, ENTREPRENEUR (Apr. 1, 2017), <https://www.entrepreneur.com/article/283990> [<https://perma.cc/Z4YY-88A9>].

248. See Kyle Vanhemert, *Simple Pictures that State-of-the-Art AI Still Can't Recognize*, WIRED (Jan. 5, 2015, 6:30 AM), <https://www.wired.com/2015/01/simple-pictures-state-art-ai-still-cant-recognize/> [<https://perma.cc/TK57-4JDR>].

249. See *id.*

250. See CHIYUAN ZHANG ET AL., UNDERSTANDING DEEP LEARNING REQUIRES RE-THINKING GENERALIZATION 3 (2017), <https://openreview.net/pdf?id=Sy8gdB9xx> [<https://perma.cc/7QQH-4QBZ>]; Robert D. Hof, *Deep Learning*, MIT TECH. REV., <https://www.technologyreview.com/s/513696/deep-learning/> [<https://perma.cc/V3BM-GQ7G>] (last visited Sept. 21, 2018); Tejani, *supra* note 246.

251. See ZHANG ET AL., *supra* note 250, at 3; Kaasschieter, *supra* note 25.

252. See Gary Marcus, *Deep Learning: A Critical Appraisal 6* (ArXiv, Working Paper No. 1801.00631, 2018) <https://arxiv.org/pdf/1801.00631.pdf> [<https://perma.cc/33YJ-LBD2>]; Jason Pontin, *Greedy, Brittle, Opaque, and Shallow: The Downsides to Deep Learning*, WIRED (Feb. 2, 2018, 8:00 AM), <https://www.wired.com/story/greedy-brittle-opaque-and-shallow-the-downsides-to-deep-learning/> [<https://perma.cc/X7Y5-J7A8>]. At its extreme, so much data may be required—e.g., quintillions of data points—that current computers would be unable to collect and process sufficient data. See Marcus, *supra*, at 7. Such constraints would, however, diminish over time as quantification of data and computing power increase. Aaron Tan, *AI Is Renewing Interest in High-Performance Computing*, COMPUTER WKLY. (Apr. 23, 2018, 8:23 AM), <https://www.computerweekly.com/news/252439705/AI-is-renewing-interest-in-high-performance-computing> [<https://perma.cc/5FUU-X54T>]. For a discussion of modern trends in data growth, see Gil Press, *IoT Mid-Year Update from IDC and Other Research Firms*, FORBES (Aug. 5, 2016, 1:41 PM), <https://www.forbes.com/sites/gilpress/2016/08/05/iot-mid-year-update-from-idc-and-other-research-firms/#4cddb58955c5> [<https://perma.cc/N7AE-UG3L>]. For a discussion of historical trends in computing power and Moore's Law, see Thomas L. Friedman, *Moore's Law Turns 50*, N.Y. TIMES (May 13, 2015), <https://www.nytimes.com/2015/05/13/opinion/thomas-friedman-moores-law-turns-50.html> [<https://perma.cc/Z2GA-DPB4>]. However, note that Moore's Law—and with it, rapid increases in computing power—have likely come to an end. Peter Bright, *Moore's Law Really Is Dead This Time*, ARS TECHNICA (Feb. 10, 2016), <https://arstechnica.com/information-technology/2016/02/moores-law-really-is-dead-this-time/> [<https://perma.cc/88SY-854V>]; Tom Simonite, *Moore's Law Is Dead. Now What?*, MIT TECH. REV. (May 13, 2016),

concept is so abstract, or open to so many interpretations that specific data either do not exist or data would point in conflicting directions, a bot could not distinguish any distinct, accurate trends in those data to guide its actions. For both reasons, high-level thought is beyond a bot's data component.

B. How Laws Operate

The struggle between specificity and ambiguity finds a legal analogue in the concepts of rules and standards.²⁵³ The former “state a determinate legal result that follows from one or more triggering facts.”²⁵⁴ The latter, in contrast, “require legal decision makers to apply a background principle or set of principles to a particularized set of facts in order to reach a legal conclusion.”²⁵⁵ For instance, a rule might state, “If you drive over sixty-five miles per hour on a highway, you will receive a \$300 ticket.” In contrast, a standard might state, “If you negligently cause harm to another, you are liable for all foreseeable damages.” Compliance with a rule is simply a matter of avoiding triggers that result in negative legal outcomes. Compliance with a standard, in contrast, requires the presence of facts that, on balance, favor one's own position, as judged by a potentially fallible decision-maker.²⁵⁶

Much scholarship has been dedicated to discussing the relative advantages, disadvantages, and applications of each.²⁵⁷ Rules are said

<https://www.technologyreview.com/s/601441/moores-law-is-dead-now-what/>
[<https://perma.cc/8YL6-EG5H>].

253. See Duncan Kennedy, *Form and Substance in Private Law Adjudication*, 89 HARV. L. REV. 1685, 1685 (1976).

254. Korobkin, *supra* note 221, at 23.

255. *Id.*

256. *Id.* To best frame the model of abstraction discussed *infra*, this Article draws a rather clear distinction between the determinate triggers of rules and the fact-intensive judgment of standards. Pierre J. Schlag, *Rules and Standards*, 33 UCLA L. REV. 379, 380 (1985). It should be noted, however, that both rules and standards may also be conceived in similar terms as legal directives, or “if this, then that” statements. *Id.* at 382–83. These directives, in turn, have two components: a “trigger” (the identifying facts that generate an outcome) and a “response” (the legal outcome). *Id.* at 381. The key difference between rules and standards lies in the types of triggers and responses: Whereas rules have a “hard empirical trigger and a hard determinate response,” standards have a “soft evaluative trigger and a soft modulated response.” *Id.* at 382–83. The response flows automatically from the trigger, regardless of what level of specificity such trigger requires. *See id.*

257. Rules and standards have been examined in a wide range of legal subfields. *See, e.g.*, Alice G. Abreu & Richard K. Greenstein, *The Rule of Law as a Law of Standards: Interpreting the Internal Revenue Code*, 64 DUKE L.J. ONLINE 53 (2015) (the Internal Revenue Code); Russell D. Covey, *Rules, Standards, Sentencing, and the Nature of Law*, 104 CALIF. L. REV. 447 (2016) (sentencing); Daniel A. Crane, *Rules Versus Standards in Antitrust Adjudication*, 64 WASH. & LEE L. REV. 49 (2007) (antitrust); John F. Duffy, *Rules and Standards on the Forefront of Patentability*,

to amplify clarity and, thus, predictability—the hallmark of the rule of law.²⁵⁸ They amplify efficiency by lowering transaction costs, as well as the ability to administrate, by simplifying the decision-making of the courts in the large numbers of cases that such courts must decide.²⁵⁹ Such clarity also leads to equal treatment, insofar as all similarly situated individuals receive the same result.²⁶⁰ Rules, however, will not achieve “perfect” results like standards may, insofar as every generalization will have instances in which it is over or underinclusive.²⁶¹

Standards, in contrast, prioritize fairness to avoid injustice, particularly the duplicity of individuals who may exploit rules to take advantage of others.²⁶² Moreover, standards may be the only option in scenarios that are too unpredictable, nuanced, or variable to be codified into a rule.²⁶³ Timing is key: Rules determine proper conduct *ex ante* at the time of legislating the rule, whereas standards do so *ex post* at the time of considering the facts.²⁶⁴

More important for constructing a set of AI Laws is understanding the specificity of rules and particularly the ambiguity of standards. As discussed, bots can grasp explicit commands in their instruction component, leaving rules a feasible option.²⁶⁵ For instance, a bot can easily be programmed not to drive over sixty-five miles per hour where that is the speed limit. However, bots have difficulty grasping extreme abstraction or ambiguity, although they can identify data-driven patterns where large amounts of data exist.²⁶⁶ The key question, then, is whether ambiguous standards can be reduced into a set of identifiable patterns.

51 WM. & MARY L. REV. 609 (2009) (patents); Edward Lee, *Rules and Standards for Cyberspace*, 77 NOTRE DAME L. REV. 1275 (2002) (cyberspace); Carol M. Rose, *Crystals and Mud in Property Law*, 40 STAN. L. REV. 577 (1988) (property law).

258. Clifford G. Holderness, *A Legal Foundation for Exchange*, 14 J. LEGAL STUD. 321, 323 (1985).

259. See Antonin Scalia, *The Rule of Law as a Law of Rules*, 56 U. CHI. L. REV. 1175, 1178 (1989) (“The common-law, discretion-conferring approach is ill suited, moreover, to a legal system in which the supreme court can review only an insignificant proportion of the decided cases.”).

260. See *id.*

261. *Id.* at 1177 (“All generalizations (including, I know, the present one) are to some degree invalid, and hence every rule of law has a few corners that do not quite fit. It follows that perfect justice can only be achieved if courts are unconstrained by such imperfect generalizations.”); see also Lin, *supra* note 176.

262. See Kennedy, *supra* note 253, at 1773.

263. See Korobkin, *supra* note 221, at 25–26.

264. See Frederick Schauer, *The Convergence of Rules and Standards*, 2003 N.Z. L. REV. 303, 305 (2003).

265. See *supra* Section IV.A.

266. See *supra* Section IV.A.

The law comes with structure and context that mitigate some of the ambiguity of legal reasoning. Parties in court must specify the applicable claims, defenses, and facts. Even where judges are tasked with determining what is “reasonable” under the circumstances (perhaps the broadest of legal standards), the judge’s discretion would be limited by the corresponding set of facts and legal arguments. Moreover, although administration of standards may have minimal precedential effect,²⁶⁷ judges may look to past cases with similar fact patterns for guidance in reaching their own decisions.²⁶⁸ Thus, standards do not exist in a vacuum of infinite vagueness.

In addition, rules and standards lie on a legal spectrum.²⁶⁹ The two can have characteristics of the other, or merge into one another.²⁷⁰ Rules may, for instance, have several exceptions that make them less predictable and, as such, are more like standards.²⁷¹ Standards, in contrast, can involve multi-factor balancing, making them more predictable and, thus, more rule-like.²⁷² Consider, for instance, a standard that states, “Do what is reasonable under the circumstances.” Compare this with a standard that states, “Do what is in the best interests of the child,”²⁷³ which explicitly requires weighing the child’s preferences, child’s age and mental development, presence of abuse or domestic violence in the household, and the parents’ mental health, among other factors. Such factors place constraints on the standard, limiting the types of facts that can be considered and thus the universe of potential options. With less options comes less ambiguity.

When considering legal ambiguity and the social risks that bots may present, criminal law bears special mention. Perhaps more than any other area of law, criminal law eschews high degrees of ambiguity,

267. Kennedy, *supra* note 253, at 1685.

268. See Schlag, *supra* note 256, at 410–11 (“A standard is open-ended only to the degree that it is stable. Standards are open-ended in that they do not create rigid boundaries and can be applied to new and unforeseen situations. Yet if the meaning or content of the standard does not remain in some sense fixed, the standard will not be open-ended at all. If there is no constancy to the standard, if it is fully open, has no boundaries, can apply to any situation, then it will not be open-ended. It will be meaningless.”).

269. Korobkin, *supra* note 221, at 24–25.

270. *Id.* at 25; see also Rose, *supra* note 257, at 580 (noting that rules (crystals) and standards (mud) oscillate from one to the other, as crystals are transformed to mud to make them more equitable, while mud is transformed back to crystals to add further clarity); Schauer, *supra* note 264, at 303.

271. See Korobkin, *supra* note 221, at 27. Korobkin’s example is of a speed limit that is excused when individuals exceed the speed limit to avoid other cars or take a sick person to the hospital. *Id.*

272. *Id.* at 28.

273. See Schauer, *supra* note 264, at 308.

particularly under the vagueness doctrine.²⁷⁴ Criminal laws are held invalid under the vagueness doctrine if they do not provide reasonable guidance to allow individuals to know what behavior is prohibited.²⁷⁵ The Due Process Clauses of the Fifth and Fourteenth Amendments of the US Constitution require that individuals be given sufficient notice of the nature of the crimes they can commit.²⁷⁶ The law cannot be arbitrarily enforced at the whim of law enforcement individuals.²⁷⁷ All individuals must be equal in the eyes of the law.²⁷⁸ Vagueness undermines that guarantee and, as such, is the enemy of enforcement.

More generally, poorly defined concepts tend to fall outside the scope of the law because they cannot be effectively enforced.²⁷⁹ Some concepts may be so abstract that they cannot be definitively defined or understood. Just as the concept of “harm” is vague to a computer, so would a generic law solely stating that “humans shall not harm one another.”²⁸⁰ Laws also avoid codifying ethical values that are open to conflicting interpretations,²⁸¹ unless a legal tradition exists that requires that laws codify one set of norms.²⁸² Highly vague or

274. See Alafair S. Burke, *Unpacking New Policing: Confessions of a Former Neighborhood District Attorney*, 78 WASH. L. REV. 985, 1018 (2003).

275. *Skilling v. United States*, 561 U.S. 358, 364 (2010) (“To satisfy due process, a ‘penal statute [must] define the criminal offense [1] with sufficient definiteness that ordinary people can understand what conduct is prohibited and [2] in a manner that does not encourage arbitrary and discriminatory enforcement.’”).

276. *Vagueness Doctrine*, CORNELL L. SCH.: LEGAL INFO. INST., <https://www.law.cornell.edu/wex/vaguenessDoctrine> [<https://perma.cc/CNN7-EHTB>] (last visited Sept. 19, 2018).

277. *Id.*

278. U.S. CONST. amend. XIV, § 1. For instance, the Equal Protection Clause of the Fourteenth Amendment provides that no state will deny any person the “equal protection of the laws.” *Id.* Internationally, Article 7 of the Universal Declaration of Human Rights by the United Nations enshrines this concept, stating that “All are equal before the law and are entitled without any discrimination to equal protection of the law.” G.A. Res. 217 (III) A, Universal Declaration of Human Rights (Dec. 10, 1948).

279. See Tim Searchinger, Note, *The Procedural Due Process Approach to Administrative Discretion: The Courts’ Inverted Analysis*, 95 YALE L.J. 1017, 1029 (1986).

280. Criminal law, in fact, places a much higher bar under the vagueness doctrine. For instance, the Supreme Court has voided statutes for penalizing “gangsters” or broad definitions of vagrancy. See *Papachristou v. City of Jacksonville*, 405 U.S. 156, 171 (1972) (regarding vagrancy statute); *Lanzetta v. New Jersey*, 306 U.S. 451, 458 (1939) (regarding statute against gangsters).

281. See Rick Garlikov, *Morality and Law*, <http://www.garlikov.com/philosophy/moralityandlaw.htm> [<https://perma.cc/SW22-SGCP>] (last visited Sept. 19, 2018). In some instances, however, a legal tradition exists that requires codifying one set of norms over another, even if some individuals disagree. For instance, the US Constitution generally prevents discrimination on the basis of race or religion, whether or not citizens wish for discriminatory legislation. See *Lanzetta*, 306 U.S. at 458.

282. See Garlikov, *supra* note 281.

multivalent moral concepts, such as the concepts found in morally focused AI Laws,²⁸³ do not typically get codified into law.²⁸⁴

To sum, the law encodes manageable ambiguity. Standards may leave substantial discretion to legal decision-makers, but that discretion still receives guidance in three forms: the facts within each case, multi-factor balancing requirements (where applicable), and legal thinking surrounding vagueness and enforceability. Rules, too, may have some standard-like exceptions that introduce ambiguity, but they favor straightforward triggers that leave little to the judicial imagination.

C. Encoding Laws into a Machine

The way laws operate parallels with the way bots operate. Just as legal rules provide clear-cut instructions regarding prohibited conduct, so too does a computer's instruction component clearly direct it to engage in or refrain from prohibited conduct. Conversely, legal standards account for ambiguity by requiring fact-specific considerations, including analysis of precedents or nonprecedential similar cases to determine the appropriate outcome on a case-by-case basis. Similarly, a computer's data component engages in a fact-specific analysis, processing each dataset it possesses and learning to act in the optimal way given its specific data.²⁸⁵ These datasets effectively serve

283. See *supra* Section II.B.

284. Garlikov, *supra* note 281. Note that many other matters also fall outside the scope of the law due to enforceability concerns. This Article emphasizes highly ambiguous moral concepts, which dovetails with the discussion of AI Laws. However, even clear or easy-to-implement moral matters are often not codified, as enforcing them would produce suboptimal outcomes. See *id.* Some would argue that smoking or drinking alcohol is immoral, but such matters are legally permitted. Regulating such conduct would place too heavy a burden on law enforcement resources such as labor and prison space. Less controversially, some matters lack sufficient social importance to regulate, particularly because liberty and autonomy demand that people be allowed to make minor mistakes. See *id.* For example, officers do not imprison temperamental bosses who yell at their employees or relatives who squabble with one another. Mandating arguably moral conduct in such instances would impose a greater social harm than the arguably immoral conduct itself. See *id.*

285. See Marr, *supra* note 217. More technically, a bot reaches an optimal result given a specific dataset, model parameters, and model hyperparameters. If any of these are not optimal (or close to optimal), the bot will accordingly also not reach an optimal result (or something close to it). The parameters refer to the variables that a machine learning model will use data to learn or estimate. Jason Brownlee, *What is the Difference Between a Parameter and a Hyperparameter?*, MACHINE LEARNING MASTERY (July 26, 2017), <https://machinelearningmastery.com/difference-between-a-parameter-and-a-hyperparameter/> [https://perma.cc/92DM-WW4B]. The hyperparameters refer to variables outside the model that cannot be estimated using data. *Id.* Frequently, the developer will specify hyperparameters. *Id.* The choice of parameters and hyperparameters can largely determine whether a machine learning model learns accurately. *Id.*

as a bot's precedent,²⁸⁶ allowing it to extrapolate and generalize to new data it has not yet processed. Finally, just as the law eschews highly abstract concepts found in different ethical systems, bots cannot process or understand high levels of abstractions. Table 2 illustrates the taxonomy that is the result of combining these legal and technical functions:

Table 2. Legal and Technical Taxonomy Summary

Legal Function	Technical Equivalent	How to Encode	Example
Rules	Low-Level Instruction	Hard-Code into a Bot	"Do Not Kill"
Standards	Data	Fact-Intensive Analysis of Data	Self-Defense
Unenforceable Matters	High-Level Thoughts	Impossible to Encode	Dignity, Goodness

This taxonomy emphasizes the major critique against Asimov's Laws of Robotics and similarly inspired AI Laws: abstraction.²⁸⁷ The Laws of Robotics try to operate in the arena of high-level thought. That, however, disregards how both the law and machines work. Some degree of clarity and guidance must be involved, both for judges to enforce laws and computers to perform assigned tasks.

Similarities between the law and computer code make it possible to translate laws into code. Legal rules are sufficiently specific and

286. One key distinction between judges and bots will likely be how dicta are handled. In theory, dicta have no precedential effect and should not be outcome-determinative in any given case. One option to train bots, then, would be for human operators to distinguish between the precedent and dicta in each case, or otherwise train a bot to make that distinction. The bot would then be trained on, and make judgments based on, only the precedential data. However, that may pose significant issues, including administrative burden and errors of interpretation. More importantly, such distinctions may be unnecessary for a bot. Given a bot's fine-tuned ability to decipher, process, and assign importance to data (including distinctions imperceptible to humans), a bot could simply treat pieces of dicta as additional data points. If the applicable data points truly did not influence a judge's decision (as would be expected, since dicta nominally has no impact on a decision), the bot would assign a low weight to such data (converging to 0, or no impact). However, if such data points did have an influence on the outcome, the bot would also realize as much and factor the applicable facts (dicta) into its decision-making. In doing so, the bot would likely increase the accuracy of its decisions—and potentially reveal that dicta may be more relevant to judges than expected.

287. See Marcus, *supra* note 77; *supra* Part II.

imperative to be directly encoded into a machine, thus prohibiting the machine from engaging in certain conduct. Legal standards present too much ambiguity to be directly encoded, but they do share two significant technical and practical characteristics with a machine's data component. On the technical front, standards and the data component handle manageable abstraction.²⁸⁸ On the practical front, each taps into an extensive store of prior data to decide ambiguous matters.²⁸⁹

The greatest technical difficulty in encoding abstraction into a computer may be distinguishing learnable and unlearnable abstractions. Within the law, however, the universe of acceptable abstraction is largely determined in the form of legal standards. Although these standards may prove ambiguous, their applications must nevertheless be based in facts. Judges must have sufficient facts to guide their judgment, even if other judges may disagree with the outcome. Past cases, moreover, provide precedent or persuasive authority to judges to ensure consistency of opinions, at least within the same jurisdiction. A bot's data component is specifically built to handle this type of fact-intensive, precedential reasoning.²⁹⁰ A bot, too, will use discernible facts from prior data to arrive at solutions for current tasks.²⁹¹ That could extend to legal compliance as bots process legal texts and past case law.

Unlike other abstract concepts, legal standards also have a straightforward solution for a bot's practical data needs: legal texts. AI requires unprecedented amounts of data.²⁹² There may be no million-point dataset to track concepts like harm, dignity, or autonomy—but there already exists a sufficiently large dataset for legal standards in the form of case law, legal treatises, and legislative history.²⁹³

Case law would prove a particularly useful source of data. The United States has a rich, voluminous, ever-growing body of case law,²⁹⁴

288. See *supra* Table 2.

289. See *id.*

290. See, e.g., NSTC PLAN, *supra* note 55, at 27; see also *supra* note 286 for a discussion of the role of dicta in such decision-making.

291. See Marr, *supra* note 217. One problematic scenario might occur where equivalent courts of a given jurisdiction disagree as to the result in nearly identical fact patterns. In such situations, a bot would simply pick the "better" outcome, depending on the nuances it has learned from examining all other case law. Doing so might result in more consistent outcomes than the decisions of various judges.

292. NSTC PLAN, *supra* note 55, at 17–18.

293. See, e.g., Westlaw, THOMAS REUTERS, <https://legalsolutions.thomsonreuters.com/law-products/westlaw-legal-research/?CID=TRSite> [<https://perma.cc/7AYT-6PHT>] (last visited Sept. 20, 2018).

294. *Access to Court Opinions Expands*, U.S. CTS. (Jan. 31, 2013), <http://www.uscourts.gov/news/2013/01/31/access-court-opinions-expands> [<https://perma.cc/5QFH-T2TM>]. There are over 600,000 federal court opinions since 2004. *Id.* There will undoubtedly be

which may be sufficient to satisfy the intensive data needs of bots. Moreover, insofar as bots of the future do behave in prohibited ways that then result in civil or criminal liability, case law may grow around AI-specific issues. That case law can then be used to further train bots in the issues most relevant to their operation.

Of course, case law and other legal data have largely not been collected or collated,²⁹⁵ and a bot would likely require significant human input in order to process such data accurately.²⁹⁶ However, bots have already shown significant progress in processing human text.²⁹⁷ Legal AI companies sift through various legal sources, including case law and legal treatises, to determine applicable case law.²⁹⁸ With human oversight, they can even provide multi-paragraph answers to legal questions posed by lawyers.²⁹⁹ The results generated by legal AI technology are already good enough that, in 2016, BakerHostetler became the first law firm to hire an AI lawyer, specifically to aid in its bankruptcy practice.³⁰⁰ With sufficiently similar analysis of textual sources, bots will likely become capable of analyzing other areas of the law as well.

D. Conclusion: The Path to the Bot Legal Code

Parts II and III demonstrate the need to eschew moral proposals of AI Laws in favor of legal and technical rules that address the full spectrum of activities that a bot will perform and commensurate risks. Part IV then identifies the parallels between the law and technical aspects of a bot, such that the law can be codified into the bot's computer

countless more state and federal opinions in the more than 200 years during which American courts have been active.

295. For examples of currently available datasets regarding case law and other legal matters, see *Data Sources: Courts*, DUKE L., <https://law.duke.edu/lib/facultyservices/empirical/links/courts/> [<https://perma.cc/4BKU-EHKJ>] (last updated April 28, 2016). These datasets unfortunately cover only a small sample of the available legal materials.

296. Note that bots have tended to perform well on supervised learning tasks (those in which a bot's training data are clearly labeled with the right answer), but not unsupervised learning. For the distinction between supervised and unsupervised learning, see Nikki Castle, *Supervised vs. Unsupervised Machine Learning*, ORACLE: DATASCIENCE.COM (July 13, 2017), <https://www.datascience.com/blog/supervised-and-unsupervised-machine-learning-algorithms> [<https://perma.cc/R6H9-ESXP>].

297. See Javier Couto, *Deep Learning for NLP, Advancements and Trends in 2017*, TRYOLABS (Dec. 12, 2017), <https://tryolabs.com/blog/2017/12/12/deep-learning-for-nlp-advancements-and-trends-in-2017/> [<https://perma.cc/2X8D-V5DV>].

298. Lohr, *supra* note 219.

299. *Id.*

300. Karen Turner, *Meet 'Ross,' the Newly Hired Legal Robot*, WASH. POST (May 16, 2016), https://www.washingtonpost.com/news/innovations/wp/2016/05/16/meet-ross-the-newly-hired-legal-robot/?utm_term=.cf89d1e9d7cc [<https://perma.cc/H34B-5TML>].

code (whether through explicit coding or the bot's own data analysis). These Parts demonstrate that sophisticated bots can, and should, comply with all applicable laws.

Having addressed these conceptual questions regarding the proper character of a set of AI Laws, the question now becomes a practical one: How do we code, develop, and implement those AI Laws? What are the necessary or ideal qualities? How do we reduce the associated administrative burden? The Bot Legal Code attempts to provide an answer.

V. THE BOT LEGAL CODE

This Article now examines in more depth what a "Bot Legal Code" should resemble and how to build it. First, this Part analyzes the ideal qualities of such a Code, which would both ensure the Code's technical feasibility and legal optimality. Then, this Part discusses how government and decentralized, peer production communities can cooperate to build and open-source the necessary software so that all bots can seamlessly develop the ability to comply with laws consistently and at minimal cost.

A. *Ideal Qualities in the Bot Legal Code*

The Bot Legal Code would contain four ideal qualities: (1) modularity, (2) depth, (3) consistency, and (4) conflict resolution. Modularity refers to solutions that are built into modules, which are compartmentalized combinations of code that can be easily transferred from one bot to another.³⁰¹ Depth requires that a bot comply with the entire gamut of laws applicable to it (or a similarly situated human agent). Consistency requires that all similarly situated bots follow the same laws in the same ways. Conflict resolution requires that a bot have a way to resolve all contradictions in its instructions. Together, these qualities ensure that any Bot Legal Code is both technically feasible and legally complete, insofar as it will accurately reflect the legal system and can be implemented seamlessly across a wide range of bots and situations.

301. *Modular Programming*, TECHOPEDIA, <https://www.techopedia.com/definition/25972/modular-programming> [<https://perma.cc/G3G9-LV4U>] (last visited Sept. 20, 2018).

1. Modularity

The Bot Legal Code must be modular so that it can be seamlessly integrated into the millions to billions of bots that will operate in the future. By building the Bot Legal Code into self-contained modules, those modules can then be implemented into larger projects as necessary.³⁰² The purpose of modularity is to maximize code reuse, thereby minimizing waste of developer time and inconsistencies within code.³⁰³

With bots, such economy of coding will be paramount for two reasons. First, there may be too many bots to manage manually. Particularly, as bots begin to write their own code or generate other bots,³⁰⁴ those bots must have consistent encoded controls to ensure that they operate properly. Second, the laws that apply to bots will vary widely from jurisdiction to jurisdiction, whether domestically or internationally. The United States contains federal laws and a separate set of laws for each of the fifty states and each territory. Similarly, many nations contain both federal and provincial-level legal systems. Ensuring that bots simultaneously comply with each of the dozens to thousands of legal systems with which they interact may not be manually possible. The technical resources necessary for each business with an AI component to do so would be unduly burdensome, if not impossible.³⁰⁵ However, such complexity can be efficiently managed by building common repositories of jurisdiction-specific modules that serve as the final definition of laws applicable in the given jurisdiction.³⁰⁶ Bots could then simply implement the modules they required.

302. See Scott Drew Pendleton et al., *Perception, Planning, Control, and Coordination for Autonomous Vehicles*, 5 MECHATRONICS: INTELLIGENT MACHINES 1, 2–3 (2017), <http://www.mdpi.com/2075-1702/5/1/6> [<https://perma.cc/PGF5-LQGJ>]. For instance, in programming a self-driving car, one could build modules to solve specific problem areas, such as perception, planning, and control, which are further divided into submodules. See *id.*

303. See Manuel Sojer & Joachim Henkel, *Code Reuse in Open Source Software Development: Quantitative Evidence, Drivers, and Impediments*, 11 J. ASS'N INFO. SYS. 868, 886–88 (2010).

304. See Dave Gershgorn, *Microsoft's AI is Learning to Write Code by Itself, Not Steal It*, QUARTZ (Mar. 1, 2017), <https://qz.com/920468/artificial-intelligence-created-by-microsoft-and-university-of-cambridge-is-learning-to-write-code-by-itself-not-steal-it/> [<https://perma.cc/G4HC-AMVL>]; Andrew Pierno, *What Happens When Bots Start Writing Code Instead of Humans*, VENTUREBEAT (Nov. 26, 2016, 2:10 PM), <https://venturebeat.com/2016/11/26/what-happens-when-bots-start-writing-code-and-web-development-dies/> [<https://perma.cc/FJ9B-H6YM>].

305. See *infra* Section V.B. But see Michael J. Garbade, *Top 8 Open Source AI Technologies in Machine Learning*, OPENSOURCE.COM (May 15, 2018), <https://opensource.com/article/18/5/top-8-open-source-ai-technologies-machine-learning> [<https://perma.cc/QG56-W4RD>].

306. See *generally* NPM, <https://www.npmjs.com/> [<https://perma.cc/LZ5Y-J6TU>] (last visited Sept. 20, 2018). This type of structure should come as no surprise to programmers, who commonly

Modularity also encourages implementation of the Bot Legal Code across a wide variety of programming languages. Notably, AI can be programmed in a variety of programming languages and frameworks,³⁰⁷ though it has shown convergence to a few popular options.³⁰⁸ For instance, the Python programming language is the most popular language for AI development³⁰⁹ and TensorFlow (built on top of Python) is the most popular framework.³¹⁰ Self-contained modules can be encoded into whatever language may be most necessary, pressing, or popular for the types of bots implementing such code, and gradually converted into other languages and frameworks as necessary.³¹¹

2. Depth

Depth concerns the content of the Bot Legal Code, which must reflect the complete set of laws that would apply to a given bot. Just as humans cannot cherry-pick which laws to follow, so should a sophisticated bot understand and follow all laws applicable to it. That may, of course, be onerous. An AGI, for instance, may need to comply with nearly all laws that would apply to humans engaging in similar activities. That is simply the price of developing sophisticated machines. However, most bots, such as manufacturing robots, will only

deal with modules. For instance, the programming language Node currently has over 700,000 open-source modules, which can be installed into any Node program. *See generally id.*

307. *See* Jerry Hamed, *The Highest Paying Programming Languages of 2016*, EDX (Sept. 14, 2016), <https://blog.edx.org/highest-paying-programming-languages-2016> [<https://perma.cc/37NS-QUHX>] (noting that there are hundreds of programming languages in existence).

308. *See* Jean Francois Puget, *The Most Popular Language for Machine Learning Is . . .*, IBM: COMMUNITY (Dec. 19, 2016), https://www.ibm.com/developerworks/community/blogs/jfp/entry/What_Language_Is_Best_For_Machine_Learning_And_Data_Science?lang=en [<https://perma.cc/89A6-YUL5>]. Note that AI is a subfield of machine learning, the topic of this Article.

309. *Python Is the Most Popular Programming Language Today for Machine Learning*, JAXENTER (Sept. 6, 2016), <https://jaxenter.com/python-popular-programming-language-today-machine-learning-128942.html> [<https://perma.cc/U78S-Y428>].

310. James Vincent, *Google's Latest Platform Play Is Artificial Intelligence, and It's Already Winning*, VERGE (May 18, 2017, 1:25 PM), <https://www.theverge.com/2017/5/18/15657256/google-ai-machine-learning-tensorflow-io-2017-platform-play> [<https://perma.cc/96D9-PTTH>].

311. *Frequently Asked Questions*, TENSORFLOW, <https://www.tensorflow.org/guide/faq> [<https://perma.cc/67FN-8L59>] (last visited Sept. 20, 2018). Converting functionality of a given program into another programming language is quite common. For instance, developers can interface with the popular artificial intelligence framework, TensorFlow, in a variety of programming languages, including both official interfaces by the TensorFlow team (in Python, C++, Java, and Go) and unofficial interfaces by the open-source community (C#, Julia, Ruby, Scala, etc.). *Id.* The TensorFlow website even offers suggestions for converting core functionality into other languages. *Tensorflow in Other Languages*, TENSORFLOW, https://www.tensorflow.org/extend/language_bindings [<https://perma.cc/M3RD-LAUZ>] (last visited Sept. 20, 2018).

need to understand a limited subset of laws, since they will only perform a limited number of activities. The Bot Legal Code can begin simply by encoding laws on an as-needed basis, but then gradually growing to keep pace with the level of sophistication of active bots.

The scope of the laws applicable to bots will, moreover, be fleshed out as bots develop and new legislation is enacted. For bots engaged in traditionally human activities, those laws will likely correspond largely to those applicable to humans engaged in similar activities.³¹² However, a bot could also have laws that uniquely apply to it, such as requiring disclosure that it is a machine³¹³ or prohibiting processing of certain types of sensitive data.³¹⁴ Not every law will apply to a mechanical agent either. Consider, for instance, labor regulations regarding salary, overtime, sick days, or vacation, which would be nonsensical if applied to a bot. The Bot Legal Code thus improves on Etzioni's first AI Rule insofar as it requires bots to comply with their own applicable laws, rather than their human operator's.

3. Consistency

Consistency requires all bots to behave similarly with respect to legal obligations. Just as the law is meant to be applied equally to all humans,³¹⁵ so should it apply equally to all bots. Legal bots must all follow the same set of laws. Their understanding of those laws, moreover, should be consistent to ensure they comply with those laws in the same manner.³¹⁶ Conversely, if each bot were to learn to follow those laws on its own, all bots may not arrive at the same, accurate understanding of legally prohibited conduct. Given the number of bots that will be active in the future, inconsistent legal training of bots could

312. See Etzioni, *supra* note 47.

313. See *id.*

314. See Rand Hindi, *Will Artificial Intelligence Be Illegal in Europe Next Year?*, *ENTREPRENEUR* (Aug. 9, 2017), <https://www.entrepreneur.com/article/298394> [<https://perma.cc/8QTJ-UQLN>]. For instance, the European General Data Protection Regulation arguably puts stringent limits on how AI (among other things) can collect and process personal information. See *id.*

315. See U.S. CONST. amend. XIV, § 1 (requiring the states to provide citizens with “equal protection of the laws”); Chester James Antieau, *Equal Protection Outside the Clause*, 40 CALIF. L. REV. 362, 362–64 (1952) (discussing the Fifth Amendment's implicit guarantee of equal protection under its due process clause).

316. Note that this quality (more than the others) expects more of the bot than would be expected of a similarly situated human. Humans do not behave consistently, nor do they agree on the interpretation of all laws. However, given the potential far-reaching capabilities of bots, we unfortunately do not have the luxury of allowing bots to take aggressive interpretations of the law, settle on “gray areas” of the law, or engage in efficient breaches or violations of law.

result in widespread illegal conduct. Consistency, in turn, requires that bots be given the same legal software or algorithm.

4. Conflict Resolution

Finally, conflict resolution requires that bots have a way to resolve the various conflicting legal instructions that they may face. Particularly, conflicts may arise in three situations: (1) among laws of various jurisdictions; (2) among laws within the same jurisdiction; and (3) internal conflicts within a given law. Western jurisprudential tradition already provides a model for addressing all three issues.

First, many bots will subscribe to the laws of multiple jurisdictions, in which case the laws of some jurisdictions may overlap with one another and conflict. For instance, in the United States, federal law trumps conflicting state law under the Supremacy Clause.³¹⁷ Thus, the Bot Legal Code would handle conflicts between federal and state law by simply following the federal law. Similarly, under the principle of *lex superior*, US constitutional law would trump both federal and state law.³¹⁸

Second, the laws of a given jurisdiction may impose conflicting requirements. In such instances, Western courts generally apply two doctrines: *lex specialis* and *lex posterior*.³¹⁹ The former requires that, when faced with a contradiction between two laws of differing degrees of specificity, the more specific law be applied.³²⁰ The latter requires that, when faced with a contradiction between two laws of equal specificity, the later-enacted law be given priority.³²¹

Finally, laws may have internal conflicts due to internal contradictions or ambiguities. For AI-based legal compliance, criminal law resolves this issue most effectively—holding laws void for vagueness where they are so unclear that a reasonable person would not understand what conduct is prohibited.³²² Just as a bot cannot process high-level ambiguity, it would also not be expected to comply with highly vague criminal laws (at least without human instruction to the contrary). More generally, when considering ambiguous statutes (including when ruling on a vagueness defense), judges apply rules of

317. U.S. CONST. art. VI, cl. 2.

318. Ralf Michaels & Joost Pauwelyn, *Conflict of Norms or Conflict of Laws?: Different Techniques in the Fragmentation of International Law*, 22 DUKE J. COMP. & INT'L L. 349, 354 (2012).

319. *See id.*

320. *Id.*

321. *Id.*

322. *See Vagueness Doctrine, supra note 276.*

statutory interpretation, such as by examining vague terms in light of their ordinary, dictionary meaning, or considering legislative intent.³²³ However, the process of judicial interpretation can itself be quite uncertain, as judges and scholars neither agree on one proper method of statutory interpretation nor apply those methods consistently.³²⁴ There are dozens of potential rules as well. This area, then, does pose more uncertainty and difficulty for AI-based processing.

A bot could implement the aforementioned conflict resolution mechanisms. Notably, the computing model of legal abstraction is already organized in order of specificity, streamlining the implementation of *lex specialis*.³²⁵ Explicit instructions, corresponding to legal rules, are by default more specific and clear-cut than matters learned through the data component, which corresponds to legal standards.³²⁶

To capture the full extent of conflict resolution doctrines would only require four modifications to the basic structure of the computing model of abstraction. First, different jurisdictions would need to be ordered in priority, such that multijurisdictional conflicts could be resolved (e.g., constitutional law would trump federal law, which would trump state law, which would trump common law). Second, the Bot Legal Code would also require the chronology of all laws within the Code, so that *lex posterior* could be applied. Third, bots would need to be trained to identify specificity, such that they could also order hard-coded instructions by specificity and resolve conflicts among various legal rules.³²⁷ Fourth, if any conflicts remained, the applicable bot would simply refuse to comply with the conflicting portions of the affected laws until provided further instruction by human operators. Although a bot could reach a decision in such situations, requiring human input in such corner cases would mitigate the risk of errors in perhaps the most difficult scenarios.

323. See LARRY M. EIG, CONG. RESEARCH SERV., STATUTORY INTERPRETATION: GENERAL PRINCIPLES AND RECENT TRENDS 2 (2011).

324. See *id.*

325. See *supra* Section IV.A.

326. See *supra* Table 2.

327. This may be one of the more difficult technical tasks to program, particularly with respect to judging the specificity of standards. One potential avenue to explore may be to use a bot's own certainty regarding its interpretation. When training on data, learning algorithms generate an error rate or metric, with more general, complex or difficult-to-understand concepts typically having higher error rates for a given amount of data (i.e., they require greater amounts of data and processing to generate similar results as more specific or simple concepts). Running a bot's learning algorithm on a fixed amount of data and comparing error rates (particularly if they are far apart) could shed light on which of two standards is the more specific.

Notably, internal conflicts regarding a statute's interpretation might pose a special difficulty. A bot could potentially be trained to apply rules of statutory interpretation, but that may not help if judges themselves disagree as to the proper results. In such instances, the most appropriate solution would likely be to simply wait for relevant case law to interpret such statutes and only then translate (whether by humans³²⁸ or the bot's own data analysis, as appropriate) such law as interpreted into the Bot Legal Code. Moreover, a bot could be trained to identify areas in which it is unusually uncertain (i.e., has an unusually high error rate) about the proper course of conduct. In such an instance, the bot could prompt its supervisors to examine the legal ambiguity, provide clarity, or direct it how to behave. The bot would simply refuse to act on such areas, unless provided further clarification or instruction by its supervisors or in an updated version of the Bot Legal Code. The bot's supervisors would, of course, need to test and choose an appropriate certainty threshold for the bot to take actions without supervision (whether 51 percent, 75 percent, 99 percent, or something else). Such a threshold could also vary depending on the level or perception of risk with respect to the applicable domain of activity.³²⁹

B. The Open-Source Solution

One option to build the Bot Legal Code would be to place the onus on the companies building and profiting from bots. In this line of thought, legal compliance is simply another cost of selling AI-related products, as it is with all businesses.

Nevertheless, as AI technology becomes more advanced, even small companies and sole persons can develop sophisticated bots. For example, in 2015, an individual created his own self-driving car software that converted an average car into an autonomous one—using a mere 2,000 lines of code, instead of the hundreds of thousands in other autonomous car technologies.³³⁰ AI startups are also tackling biometric

328. To mitigate issues that humans themselves may introduce into the bot by misinterpreting laws, the Bot Legal Code would likely need input from both lawyers and coders working in unison. Updates would also be necessary if, for instance, courts subsequently adopted a different interpretation of a particular law.

329. Note that the decision of an appropriate threshold would carry significant consequences for bot-based legal compliance and would itself potentially be controversial.

330. Brett Williams, *Comma.ai Just Released an \$88 Car Hacking Tool*, MASHABLE (July 17, 2017), <https://mashable.com/2017/07/07/comma-ai-launches-universal-car-interface/#gsNHwoYfRgqh> [https://perma.cc/EH8V-NRP2].

security,³³¹ forecasting the weather,³³² and interpreting medical images.³³³ With improvements in computing power and AI technology, such accomplishments will only become easier. Small to medium-sized AI enterprises, however, will likely lack the necessary legal resources to implement a complicated Bot Legal Code. Few developers are trained lawyers. Many startups, even ones with significant funding, may also choose to ignore the need for legal compliance in the race to build the most advanced bots.

Placing the burden of the Bot Legal Code on each company would also fail to be optimal. Given the thousands of AI companies in existence,³³⁴ requiring each to implement its own bespoke solution would lead to unnecessary mistakes, inconsistencies, and redundancies. Not all such companies could be monitored, nor would each follow the same precautions or have the same level of technical and legal expertise. The Bot Legal Code should be consistent and seamless.

That, in turn, argues in favor of building the Bot Legal Code as open-source software,³³⁵ or software that is freely licensed to any member of the public for personal or commercial use, redistribution, and modification without permission.³³⁶ Open-source software effectively serves as a central, shareable repository of code, which can then easily be integrated into any other computer program. Within the AI field, open-source software has played a particularly important role in disseminating access to advanced AI tools and software.³³⁷ To date,

331. See, e.g., *Facial Biometrics Authentication*, TRUST STAMP, <https://truststamp.us> [<https://perma.cc/5Y5X-LM27>] (last visited Sept. 21, 2018).

332. See, e.g., TEMPOQUEST, <http://tempoquest.com/> [<https://perma.cc/G72D-F49K>] (last visited Sept. 21, 2018).

333. See, e.g., *Transforming Patient Care with the Power of AI*, ZEBRA MED. VISION, <https://www.zebra-med.com/> [<https://perma.cc/V6Z7-6M93>] (last visited Sept. 21, 2018).

334. Currently, AngelList, a popular fundraising website for startups, lists over 5,208 AI startups. *Artificial Intelligence Startups*, ANGELLIST, <https://angel.co/artificial-intelligence> [<https://perma.cc/VFA6-8XTJ>] (last visited Sept. 21, 2018). There may, of course, be large numbers of small AI companies not listed there and many larger, established companies developing AI technologies.

335. Note that bot developers would need to program their own (potentially proprietary) software to implement this open-source software, depending on the particular activities that a bot would perform.

336. See *What is Open Source Software?*, LINUX FOUND. (Feb. 14, 2017), <https://www.linuxfoundation.org/blog/what-is-open-source-software/> [<https://perma.cc/6MCS-Q8JD>] (“Open Source software . . . is software distributed under a license that meets certain criteria: 1. It is available in source code form (without charge or at cost)[,] 2. Open Source may be modified and redistributed without additional permission[,]. 3. Finally, other criteria may apply to its use and redistribution.”).

337. Patrick Shafto, *Why Big Tech Companies Are Open-Sourcing Their AI Systems*, CONVERSATION (Feb. 22, 2016, 6:15 AM), <https://theconversation.com/why-big-tech-companies-are-open-sourcing-their-ai-systems-54437> [<https://perma.cc/JXB9-JKUF>]; see also Steven J. Vaughan-Nichols, *Microsoft Says AI and Machine Learning Driven by Open Source and the Cloud*,

many of the most popular AI-related tools and frameworks come from leading AI academics and companies, such as Google-supported TensorFlow,³³⁸ Berkeley-supported Caffe,³³⁹ academia-supported scikit-learn,³⁴⁰ and Microsoft-supported Bot Framework.³⁴¹

Moreover, open-source software can be retooled as necessary, particularly by the people who will use the code. If a particular module is incomplete or incorrect, the peer community or the bot's developers could troubleshoot and add any code that may be necessary for the bot to perform as necessary.³⁴² For instance, assume a bot transports dangerous goods like dynamite, but its developers only have access to a module regarding transport of flammable materials. Those developers could use the latter module as a starting point for building in the necessary functionality, which would likely be a much less burdensome task than building an entire machine-interpretable set of laws by themselves. For similar reasons, even large enterprises use large quantities of open-source code among their own proprietary systems.³⁴³

Of course, such capabilities also bring risks, including the risk of bad coding. A bot's developers could perceive a problem that does not exist and accidentally alter the code in a way that causes the bot to malfunction, particularly if they do not understand the Bot Legal Code on a technical level. If developers begin tweaking the Bot Legal Code, then bots may have different interpretations of the laws, reducing their predictability. However, technical mistakes are bound to happen given the complexity of coding.³⁴⁴ To the extent that the Bot Legal Code provides a basic set of coded laws—vetted by the programming

ZDNET (Mar. 8, 2018, 3:20 PM), <https://www.zdnet.com/article/microsoft-says-ai-and-machine-learning-driven-by-open-source-and-the-cloud/> [<https://perma.cc/JZ8E-GP2R>].

338. TENSORFLOW, <https://www.tensorflow.org/> [<https://perma.cc/D99E-FTA9>] (last visited Sept. 21, 2018).

339. Caffe, BERKELEY VISION, <http://caffe.berkeleyvision.org/> [<https://perma.cc/3S78-J7ED>] (last visited Sept. 21, 2018).

340. See *About Us*, SCIKIT-LEARN, <http://scikit-learn.org/stable/about.html#funding> [<https://perma.cc/RKW9-QERR>] (last visited Sept. 21, 2018).

341. Bot Framework, MICROSOFT, <https://dev.botframework.com/> [<https://perma.cc/N9GM-CU65>] (last visited Sept. 21, 2018).

342. Ryan Francis, *Open Source Security Is Not As Big of a Concern As It Once Was*, CSO (Feb. 18, 2016, 10:24 AM), <https://www.csoonline.com/article/3033837/security/open-source-security-is-not-as-big-of-a-concern-as-it-once-was.html> [<https://perma.cc/HMF6-B56U>] (“[T]he [open-source] user gets to be involved in the development process through the creation of additional features, bug reports, and code review of the projects”).

343. Saurabh Tyagi, *How Fortune 500 Is Embracing Open-Source Technology*, OPENSOURCE (July 29, 2016), <https://opensourceforu.com/2016/07/fortune-500-embracing-open-source-technology/> [<https://perma.cc/8ZPJ-MF5Y>].

344. See James Somers, *The Coming Software Apocalypse*, ATLANTIC (Sept. 26, 2017), <https://www.theatlantic.com/technology/archive/2017/09/saving-the-world-from-code/540393/> [<https://perma.cc/9AH7-CKFR>].

community—that programmers will generally not need to edit significantly, it can mitigate the many technical errors that would result were individual programmers to program their own technical compliance into their bots.

A more significant risk of open-source software may be the risk of malicious actors. A malicious actor could also study every piece of the Bot Legal Code looking for security vulnerabilities that could be used to disrupt bots that deployed the Bot Legal Code. Specifically, by increasing the visibility of source code, open-source may increase security risks, resulting in quicker and more severe exploits and increasing the number of attempted attacks.³⁴⁵

However, with highly intricate software like the Bot Legal Code, open-source may in fact reduce these forms of risks for multiple reasons. First, since open-source software is available to anyone, it can also be vetted by anyone, leading to quicker identification and correction of security vulnerabilities.³⁴⁶ Second, open-source software can be patched immediately to address any vulnerability, whereas the users of proprietary software may not have the right—let alone access to the code—to implement such fixes on their own.³⁴⁷ Finally, the Bot Legal Code would be deployed in an age of sophisticated bots, likely including bots that write computer code, identify security vulnerabilities, and potentially even patch their own vulnerabilities.³⁴⁸ These bots will have abilities far exceeding a human developer's in preemptively addressing software risks.³⁴⁹

345. See SAM RANSBOTHAM, AN EMPIRICAL ANALYSIS OF EXPLOITATION ATTEMPTS BASED ON VULNERABILITIES IN OPEN SOURCE SOFTWARE 19 (2010), https://www.econinfosec.org/archive/weis2010/papers/session6/weis2010_ransbotham.pdf [<https://perma.cc/U9MR-YUSS>] (“My theoretical development and empirical results indicate that, compared with closed source software, vulnerabilities in open source software: (a) have increased risk of exploitation, (b) diffuse sooner and with higher total penetration, and (c) increase the volume of exploitation attempts.”).

346. RUSSELL CLARKE, DAVID DORWIN & ROB NASH, IS OPEN SOURCE SOFTWARE MORE SECURE? 26 (1999), [https://courses.cs.washington.edu/courses/csep590/05au/whitepaper_turnin/oss\(10\).pdf](https://courses.cs.washington.edu/courses/csep590/05au/whitepaper_turnin/oss(10).pdf) [<https://perma.cc/M3UE-P2HQ>].

347. Ashish Arora et al., *An Empirical Analysis of Software Vendors' Patch Release Behavior: Impact of Vulnerability Disclosure*, 21 J. INFO. SYS. RES. 115, 117 (2010). Open-source communities do, in fact, tend to patch software faster than closed-source software vendors. *Id.*

348. See Rosenberg, *supra* note 13; Tom Simonite, *AI Software Learns to Make AI Software*, MIT TECH. REV. (Jan. 18, 2017), <https://www.technologyreview.com/s/603381/ai-software-learns-to-make-ai-software/> [<https://perma.cc/6MA4-DBVX>]. Some progress has been made in this area already. See, e.g., Blair Hanley Frank, *Oracle's New Database Uses Machine Learning to Automate Administration*, VENTUREBEAT (Oct. 1, 2017, 7:13 PM), <https://venturebeat.com/2017/10/01/oracles-new-database-uses-machine-learning-to-automate-administration/> [<https://perma.cc/9DFU-LJ78>].

349. See Rosenberg, *supra* note 13.

C. Government and Peer Production

Although any group or entity can distribute code as open-source, open-source development frequently occurs through peer production, a decentralized form of production through internet-connected groups of individuals.³⁵⁰ Peer production “relies on decentralized information gathering and exchange to reduce the uncertainty of participants.”³⁵¹ By utilizing the numerous skillsets of varied groups of individuals, peer production allows individuals to self-identify and work on tasks to which they are particularly suited.³⁵² As a result, “human creativity”³⁵³ tends to be allocated more efficiently in peer production.³⁵⁴ Although participants in peer production communities frequently lack monetary incentives to participate, those communities may offer sufficient alternative incentives in the form of increased reputation, better job prospects, or collective identity,³⁵⁵ among others. Peer production has resulted in remarkably accurate projects, which may perform even better than paid options.³⁵⁶ Some notable successes include “free culture projects” like Wikipedia,³⁵⁷ as well as millions of pieces of open-source software.³⁵⁸

One potential solution, then, is to simply encourage the open-source community to develop solutions for itself. If AI regulations mandated that bots must directly comply with all applicable laws, for instance, AI developers would have a common incentive to develop legally compliant software. In some instances, the open-source community may even partner with institutions that have similar interests. For example, TensorFlow currently has approximately 1,600 contributors, including many who are not affiliated with Google.³⁵⁹

350. See Yochai Benkler et al., *Peer Production: A Form of Collective Intelligence*, in HANDBOOK OF COLLECTIVE INTELLIGENCE 175, 177 (Thomas Malone & Michael Bernstein eds., 2015).

351. Yochai Benkler, *Coase's Penguin, or, Linux and the Nature of the Firm*, 112 YALE L.J. 369, 375 (2002).

352. *Id.* at 375–76.

353. *Id.* at 375.

354. *Id.* at 375–76.

355. Benkler et al., *supra* note 350, at 184.

356. *See id.* at 189.

357. *Id.* at 177.

358. *See id.* Notably, the website GitHub, which serves as a repository for software projects, contains over 5 million open-source projects. *The State of the Octoverse 2017*, GITHUB, <https://octoverse.github.com/> [<https://perma.cc/L5ZS-9XVN>] (last visited Sept. 21, 2018) (estimate calculated using programming language-specific breakdown). Most such projects do not have large sponsors.

359. GITHUB, <https://github.com/tensorflow/tensorflow> [<https://perma.cc/8ZMS-N5ED>] (last visited Sept. 21, 2018) (1,647 contributors as of date of access).

Large-scale peer production, however, tends to be uncommon,³⁶⁰ and the significant resource requirements of the Bot Legal Code may prove unduly burdensome to allow feasible peer production. The Bot Legal Code requires coding significant bodies of law (federal, state, and territorial) across many jurisdictions, domestically and internationally. Moreover, that coding will require both programming and legal expertise. Unlike programmers, who have collectively collaborated on millions of projects, lawyers do not generally collaborate on decentralized software development. Law firms are, in fact, highly protective of their proprietary documents and work product.³⁶¹ Finding altruistically minded lawyers across a variety of areas of law to provide countless hours of labor poring over statutes and legal doctrines may also be difficult. Lawyers could potentially be incentivized if such work were deemed pro bono under state bar regulations—and such work may very well be so deemed, since it would require the provision of significant legal expertise at no cost.³⁶² Even so, barring significant cooperation among large numbers of nontechnical lawyers and AI developers, such peer production will likely prove infeasible.

Rather, government production may prove a more fruitful option—leveraging greater legal, technical, and financial resources. The US government is already an exemplar of open-source activity, as it has developed thousands of open-source projects,³⁶³ has a hub for releasing such software at Code.gov,³⁶⁴ and has an open-source software policy for software developed internally and by contractors.³⁶⁵

360. See Benkler et al., *supra* note 350, at 192.

361. See, e.g., *Hickman v. Taylor*, 329 U.S. 495, 500–03 (1947). There have, however, been notable exceptions, particularly in the area of emerging technology law. Several law firms, such as Orrick, Herrington & Sutcliffe LLP and Cooley LLP, have released public templates for startup incorporation, intellectual property protection, and some financing activities, among other documents. See, e.g., *Startup Forms Library*, ORRICK, <https://www.orrick.com/Total-Access/Tool-Kit/Start-Up-Forms> [<https://perma.cc/44GN-RX6A>].

362. Model Rule 6.1 of the Model Rules of Professional Conduct calls for participation of lawyers in a wide variety of pro bono activities, including participation in activities for improving the law, the legal system, or the legal profession. See MODEL RULES OF PROF'L CONDUCT r. 6.1 (2016). Codifying a set of laws relevant to a new technology could very well be deemed to improve the law and legal system. See *id.*

363. *Sharing America's Code*, CODE.GOV, <https://www.code.gov> [<https://perma.cc/VQ9N-HPAG>] (last visited Sept. 21, 2018). Even the US Department of Defense has developed open-source software, including to renovate its internal software. See Press Release, U.S. Department of Defense, DoD Announces the Launch of "Code.mil," an Experiment in Open Source (Feb. 23, 2017); Chris Nimmer, *How Forge.mil Changed the Way the US DoD Develops Software*, OPENSOURCE.COM (Apr. 17, 2015), <https://opensource.com/government/15/4/how-forgemil-changed-way-dod-develops-software> [<https://perma.cc/752A-9QTR>].

364. *Sharing America's Code*, *supra* note 363.

365. See David A. Wheeler, *Publicly Releasing Open Source Software Developed for the U.S. Government*, CYBER SEC. & INFO. SYS. INFO. ANALYSIS CTR. (Mar. 11, 2016),

Moreover, it invests heavily in the AI space,³⁶⁶ with expenditures for unclassified AI research and development reaching \$1.1 billion in 2015 alone.³⁶⁷ Finally, it is the largest domestic provider of data, supplying over 300,000 datasets on a wide range of topics from demographic data to data on the climate, education, product recalls, and police activity.³⁶⁸ The federal government thus has expertise in perhaps the most significant area of AI development—data collection and processing. By focusing its data-related efforts on compiling the millions of legal statutes and cases, the federal government can develop the datasets necessary to train bots to understand legal standards. Thus, the federal government's software and data capabilities likely allow it to competently handle both a bot's instruction and data components.

As the largest single domestic employer,³⁶⁹ the federal government also already has the sheer amount of labor required to develop a domestic Bot Legal Code, or can otherwise scale to meet any needs. The Bot Legal Code will be a long-term, continuous project—taking several years to finish. Even after completion, the Bot Legal Code will require updates as laws change. It will also require collaboration among AI developers, data specialists, and attorneys with subject matter expertise. Fortunately, the federal government—and, to a lesser extent, state governments—hires for all such positions and can hire additional labor as necessary.³⁷⁰ Coordination may also be easier than in a peer production scenario, since the federal government has experience in employing individuals with the complete set of necessary skills, including lawyers and engineers, who do not normally collaborate on such projects.

<https://www.csiac.org/journal-article/publicly-releasing-open-source-software-developed-for-the-u-s-government/> [<https://perma.cc/D8LD-7Q27>].

366. NSTC PLAN, *supra* note 55, at 3 (“[T]he U.S. government has invested in AI research for many years.”).

367. *Id.* at 6.

368. DATA.GOV, <https://www.data.gov> [<https://perma.cc/QLA5-FULN>] (last visited Sept. 21, 2018).

369. Jennifer Calfas, *The 6 Biggest Employers in the U.S. Right Now*, TIME (Apr. 27, 2017), <http://time.com/money/4754123/biggest-us-companies> [<https://perma.cc/9T3E-M2ME>]. The Department of Defense is the single largest employer in the world, with 3.2 million employees. Niall McCarthy, *The World's Biggest Employers [Infographic]*, FORBES (June 23, 2015, 8:20 AM), <https://www.forbes.com/sites/niallmccarthy/2015/06/23/the-worlds-biggest-employers-infographic/#3127bbac686b> [<https://perma.cc/X5ZL-S4GU>].

370. See generally Beth Cobert, *Strengthening the Federal Cybersecurity Workforce*, OPM.GOV (July 12, 2016, 10:00 AM), <https://www.opm.gov/blogs/Director/2016/7/12/Strengthening-the-Federal-Cybersecurity-Workforce/> [<https://perma.cc/A6HL-5EGP>]. This, of course, assumes that politicians can agree on the importance of such research and allocate necessary resources to meet such hiring needs—unfortunately, a significant assumption.

D. Open-Source Software: Implications for the Bot Legal Code

Open-source software would satisfy all four criteria for the Bot Legal Code: depth, conflict resolution, consistency, and modularity. First, the Bot Legal Code could more easily achieve the necessary depth of coding and conflict resolution, insofar as it concentrated resources of many individuals and government actors. Lawyers with expertise in particular areas of the law, for example, would work with programmers to encode those areas of the law. Lawyers skilled in conflicts of laws could also aid with conflict resolution.

The Bot Legal Code would also ensure consistency, since all bots would follow the same Code, resulting in similar conduct in similar scenarios. That consistency would grant researchers a better opportunity to study the implementation of the Code across many bots, activities, and environments. The implications of the Code would thus become better understood over time, mitigating the unpredictability of AI. Revisions to the law and corresponding upgrades or bug fixes to the Bot Legal Code could also be rolled out simultaneously to all bots.

Finally, any such software would be modular insofar as any necessary functionality could be transferred or downloaded into any bot,³⁷¹ with minimal integration time or cost. That modularity would be best served by developing modules corresponding to separate jurisdictions or bodies of law. For instance, a self-driving car operating nationwide would subscribe to the modules corresponding to the applicable self-driving car laws in each of the fifty states and under federal law. In contrast, an AGI operating solely in California would subscribe to the modules corresponding to all California and federal law. Ideally, international modules would also be developed in other nations, allowing bots to operate across borders simply by switching to the laws of the applicable jurisdiction.

Developed modules could also be converted into whatever form necessary for a specific bot. As noted, AI programming has converged to a few programming languages and frameworks,³⁷² such as Python, the most popular programming language.³⁷³ The Bot Legal Code could first be encoded into one of the most popular languages or frameworks,

371. Often, downloading such functionality into a larger project is as simple as typing one line of code into a computer's command line. For instance, installing a Python package using Python's package manager, pip, requires typing "pip install [Project Name]." See Python Packaging Authority, *Installing Packages*, PYTHON SOFTWARE FOUND., <https://packaging.python.org/tutorials/installing-packages> [https://perma.cc/84SN-PNL2] (last visited Sept. 21, 2018).

372. See Puget, *supra* note 308.

373. *Python Is the Most Popular Programming Language Today for Machine Learning*, *supra* note 309.

which would serve as the definitive source of truth or model for the Code. The Code could then be converted into other languages or frameworks either by government actors or with help from the open-source community.

Modularity offers one further, significant benefit: the Bot Legal Code can be built gradually over time. Codifying an entire system of law will be no easy task, nor will it be necessary until and if true AGIs are developed that can do nearly any activity. Rather, bots will begin to engage in sensitive activities, gradually including driving cars (expected around 2020)³⁷⁴, acting as retail salespeople (expected around 2030),³⁷⁵ and working as surgeons (expected around 2050).³⁷⁶ As AI technologies develop, so can the Bot Legal Code to address the legal issues raised by those innovations.

VI. CONCLUSION

This Article provides a practical roadmap for developing the Bot Legal Code. Bots of the future may be uncontrollable, since they will potentially be nearly everywhere, do nearly everything, cooperate in large networks, and have access to beyond-human intelligence. Rather than futilely attempt to micromanage these intelligent machines on an ad hoc basis, we likely need to imbue them with the capability to comply with the legal systems in which they operate. Agency brings with it responsibility. As machines assume more agency, we must mitigate the risks of that agency.

Given the sophistication of bots, legal learning is likely within their grasp if the necessary resources can be collected. This Article demonstrates that technical factors are not the bottleneck to legal compliance. The computing model of legal abstraction emphasizes how legal rules and standards find their parallels in explicit coding and data-based learning. Although the law does have its ambiguities, such ambiguities tend to be grounded in the facts of each case and thus manageable. Rather, the Bot Legal Code is daunting for practical reasons. Lawyers with subject matter expertise, data specialists, and developers must collaborate across a variety of jurisdictions and areas of law. The Code must also be maintained and evolve with the law over

374. Bill Howard, *How Self-Driving Cars Work, and When They'll Get Real*, EXTREMETECH (July 11, 2017, 11:35 AM), <https://www.extremetech.com/extreme/252112-what-is-a-self-driving-car> [<https://perma.cc/3EQC-MDWY>].

375. Emerging Tech. from the arXiv, *Experts Predict When Artificial Intelligence Will Exceed Human Performance*, MIT TECH. REV. (May 31, 2017), <https://www.technologyreview.com/s/607970/experts-predict-when-artificial-intelligence-will-exceed-human-performance> [<https://perma.cc/H85A-MYGX>].

376. *Id.*

time. Peer production may provide a start, but such a Herculean task will likely require government collaboration to bring together and manage the sheer amount of labor, time, and expertise required. Government actors must be proactive, taking on such a task before intelligent bots begin to pose unmanageable risks.

