California Western Law Review

Volume 55 | Number 1 Article 4

12-20-2018

Law Without Mind: AI, Ethics, and Jurisprudence

Joshua P. Davis

LAW WITHOUT MIND: AI, ETHICS, AND JURISPRUDENCE

JOSHUA P. DAVIS*

Anything we can conceive that computers may do, it seems that they end up doing and that they end up doing it better than us and much sooner than we expected. They have gone from calculating mathematics for us to creating and maintaining our social networks to serving as our personal assistants. We are told they may soon become our friends and make life and death decisions driving our cars. Perhaps they will also take over interpreting our laws. It is not that hard to conceive of computers doing so to the extent legal interpretation involves mere description or prediction. It is much harder to conceive of computers making substantive moral judgments. So, the ultimate bulwark against ceding legal interpretation to computers—from having computers usurp the responsibility and authority of attorneys, citizens, and even judges—may be to recognize the role moral judgment plays in saying what the law is. That possibility connects the cutting edge with the traditional. The central dispute in jurisprudence for the past half century or more has been about the role of morality in legal interpretation. Suddenly, that dispute has great currency and urgency. Jurisprudence may help us to clarify and circumscribe the role of computers in our legal system. And contemplating AI may help us resolve jurisprudential debates that have vexed us for decades.

* Professor and Director, Center for Law and Ethics, University of San Francisco School of Law. I am grateful for excellent support for my research from Suzanna Mawhinney, one of our many talented research librarians, and Javkhlan Enkhbayar, one of our many talented students. My thinking on this topic benefited greatly from discussions with Brad Wendel. As always, all errors remain my own.

TABLE OF CONTENTS

I. INTRODUCTION

In WORLD WITHOUT MIND, Franklin Foer paints a bleak picture of the ways in which technological advances threaten our culture, our individuality, and, ultimately, our minds.[1] Elon Musk speculates that artificial intelligence ("AI") will likely be the cause of World War III.[2] Pauline Kim suggests that a corporation may engage in illegal employment discrimination without anyone ever knowing—or perhaps ever being able to know—that it is doing so because of reliance on AI that is ever-evolving and leaves no record of the basis for its recommendations.[3] As Jacob Weisberg writes, "Algorithms are developing their capabilities to regulate humans faster than humans are figuring out how to regulate algorithms."[4] Technological advances pose a grave threat to human ethics. Those advances also cast light on longstanding jurisprudential controversies.

Consider how puzzles presented by developing technologies give novel salience to one of the oldest and most fundamental disagreements in jurisprudence: morality's role in determining what the law is. That jurisprudential disagreement is between legal positivists and non-

1. FRANKLIN FOER, WORLD WITHOUT MIND: THE EXISTENTIAL THREAT OF BIG TECH (2017). Foer's book is one among a recent wave viewing modern technology with great skepticism and concern. *See, e.g.*, VIRGINIA EUBANKS, AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR (2018); SAFIYA UMOJA NOBLE, ALGORITHMS OF OPPRESSION: HOW SEARCH ENGINES REINFORCE RACISM (2018); NOAM COHEN, THE KNOW-IT-ALLS: THE RISE OF SILICON VALLEY AS A POLITICAL POWERHOUSE AND SOCIAL WRECKING BALL (2017); CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY 204 (2016). For an alternative approach that focuses less on risks see MAX TEGMARK, LIFE 3.0: BEING HUMAN IN THE AGE OF ARTIFICIAL INTELLIGENCE (2017); NICK BOSTROM, SUPERINTELLIGENCE: PATHS, DANGERS, STRATEGIES (2014).

2. Seth Fiegerman, *Elon Musk Predicts Word War III*, CNN (Sept. 5, 2017, 10:38 AM), https://money.cnn.com/2017/09/04/technology/culture/elon-musk-ai-world-war/index.html. After dismissing concerns about the threat of North Kora as an "existential threat" to civilization, he Tweeted, "Competition for AI superiority at national level likely cause of WW3 imo." *Id.*

3. *See* Pauline Kim, *Data-Driven Discrimination at Work*, 58 WM. & MARY L. REV. 857, 902-09 (2017).

4. Jacob Weisberg, *The Digital Poorhouse*, N.Y. REV. BOOKS (June 7, 2018), at 3, https://www.nybooks.com/articles/2018/06/07/algorithms-digital-poorhouse/?printpage=true.

positivists (sometimes called natural lawyers).[5] To oversimplify, legal positivists claim that one can say what the law is without making moral judgments about what it should be.[6] In contrast, non-positivists—at least according to this formulation—hold that moral judgments are necessary to determine what the law is.[7] This debate between positivists and non-positivists has been the focus of jurisprudence for the past fifty years.[8] Today, as we attempt to resolve the dilemmas created by technological advances, we find ourselves facing age-old issues: what role should the law play in guiding our actions and, if we have a moral obligation to abide by the law, must we consult morality in determining its content?

Autonomous cars provide a compelling example. Programming them involves not just technical knowledge, it would seem, but also moral philosophy. Imagine that an autonomous car with a single occupant is about to collide with a group of pedestrians. Furthermore, imagine that crashing into the pedestrians will kill many of them but minimize the risk to the single occupant. Alternatively, the car could swerve off a cliff, killing its occupant but sparing the pedestrians' lives.[9]

---

5. *See* Joshua P. Davis, *Legality, Morality, Duality*, 2014 UTAH L. REV. 55, 61-63 (2014). To be more precise, many positivists subscribe to what is sometimes call the "Social Fact Thesis: legal positivism holds that 'all legal facts are ultimately determined by social facts alone.'" *Id.* at 61 (quoting SCOTT J. SHAPIRO, LEGALITY 27 (2011)).

6. *Id.* at 61-63.

7. *Id.*

8. *See, e.g.*, Scott Hershovitz, *The End of Jurisprudence*, 124 YALE L.J. 1160, 1162 (2015) ("For more than forty years, jurisprudence has been dominated by the Hart-Dworkin debate."). The famous debate featured Hart L.A. Hart and Ronald Dworkin. *Id.* Hart and Dworkin's disagreement was germinal of ongoing disputes about legal positivism, with Hart likely the most influential positivist of the past half century and Dworkin the most influential non-positivist, although few current theorists may accept the view of either in an unqualified form. *See* Scott Shapiro, *The "Hart-Dworkin" Debate: A Short Guide for the Perplexed*, *in* RONALD DWORKIN 22 (Arthur Ripstein ed., 2007) ("For the past four decades, Anglo-American legal philosophy has been preoccupied—some might say obsessed—with something called the 'Hart-Dworkin' debate.").

9. Dilemmas of this sort are often called "trolley problems," a term coined by the philosopher Judith Thomson building in part on the work of Philippa Foot. *See* Bryan Casey, *Amoral Machines, or: How Roboticists Can Learn to Stop Worrying and Love the Law*, 111 NW. U. L. REV. 1347, 1353 n.36 (2017) (discussing Judith

For the autonomous car making that decision, there are at least two daunting challenges. The first is technological. The car must be able to assess the consequences of different actions available, a requirement different from—and seemingly even more challenging than—operating in a manner consistent with the ordinary rules of the road. It is no mean feat to build a car that can stop at red lights, make legal turns, avoid other vehicles, drive within the speed limit, defer to pedestrians in crosswalks, and the like.[10] Asking the car to assess the likely loss of life in two or more bad options seems significantly harder yet. But let us assume that car designers overcome that technical task—that they figure out a way for the self-driving car to determine, in a probabilistic manner, the potential adverse results in different, undesirable scenarios. A different challenge still remains.

The second challenge involves prescription, not just description or prediction. What *should* the car do? Should the car sacrifice the life of its "driver" to protect others? Does the answer depend on how many others would be at risk? Does it depend on who, if anyone, is morally responsible for bringing about the unfortunate situation in which harm must befall someone?[11]

Jarvis Thomson, *Killing, Letting Die, and the Trolley Problem*, 99 MONIST 204, 206 (1976) and Philippa Foot, *The Problem of Abortion and the Doctrine of Double Effect*, 5 OXFORD REV. 1, 3 (1967)).

10. The recent case of an Uber self-driving car killing a pedestrian provides an example of the difficulty of this task. *See* Ian Wren, *Uber Won't Seek California Permit Renewal to Test Self-Driving Vehicles After Fatal Crash*, NPR (Mar. 27, 2018, 2:49 PM), https://www.npr.org/sections/thetwo-way/2018/03/27/597331608/arizona-suspends-ubers-self-driving-vehicle-testing-after-fatal-crash (describing self-driving technology causing the tragic death of a 49-year-old woman). Of course, a single fatality does not establish that self-driving cars are less safe than human drivers. Note also that Uber's self-driving cars may also have had more difficulties than the self-driving cars of other manufacturers. *See* Daisuke Wakabayashi, *Uber's Self-Driving Cars Were Struggling Before Arizona Crash*, N.Y. TIMES (Mar. 23, 2018), https://www.nytimes.com/2018/03/23/technology/uber-self-driving-cars-arizona.html.

11. As noted above, these issues are often called "trolley problems." Casey, *supra* note 9, at 1353. There is a large literature analyzing them and conducting empirical research to see how people respond to them, and many commentators have recognized their application to self-driving cars and other new technological developments. *See, e.g.*, W. Bradley Wendel, *Economic Rationality and Ethical Values in Design-Defect Analysis: The Trolley Problem and Autonomous Vehicles*, 55 CAL. W. L. REV. 129 (2018); Casey, *supra* note 9, at 1353 n.36 (discussing Thomson, *supra* note 9, at 206, and Foot, *supra* note 9, at 1, 3).

Enter the law and lawyers. The programmers of computers in self-driving cars—and those who employ the programmers—are likely to take the law into account. In part, the relevant concern will be prudential. They will want to adjust their behavior in light of potential legal liability. Will manufacturers of cars be held liable if they instruct cars to sacrifice—or not to sacrifice—drivers under some circumstances? Will courts permit car manufacturers to give consumers options? If so, will consumers be held liable if they purchase "selfish" cars as opposed to "altruistic" ones? Will manufacturers need to disclose that risk? No doubt numerous other weighty prudential questions will arise, ones that need to be addressed for manufacturers and their customers to make informed, self-interested decisions.

Here, again, technology may have a role to play. There may come a day—perhaps sooner than we anticipate or like—when technology can engage in legal interpretation.[12] It may be able to do so as well as—perhaps better than—human beings. There was a time when many people thought a computer could never beat the best human chess player. The game is just too complicated. Then a computer did.[13] The same recently happened in Go, an ancient game, far more complex than chess, in which human beings were believed to have an even greater advantage over computers than in chess.[14] Perhaps it is only a matter of time before the same is true for legal interpretation—or at least for predicting how the law will be interpreted.

12. *See, e.g.*, RICHARD SUSSKIND & DANIEL SUSSKIND, THE FUTURE OF THE PROFESSIONS: HOW TECHNOLOGY WILL TRANSFORM THE WORK OF HUMAN EXPERTS 283 (2015) ("As machines become increasingly capable, in response to the question 'What will be left for human professionals to do?', it is also hard to resist the conclusion that the answer must be, 'less and less'."); *see also id.* at 66-71 (summarizing, inter alia, the ways in which technology can now perform tasks historically reserved to attorneys).

13. TEGMARK, *supra* note 1, at 51.

14. *World's Best Go Player Flummoxed by Google's 'Godlike' AlphaGo* AI, GUARDIAN (May 23, 2017), https://www.theguardian.com/technology/2017/may/23/alphago-google-ai-beats-ke-jie-china-go. The best Go player in the world, Ke Jie, was apparently stunned to lose to a computer, something he claimed would never happen. *Id.* His response was effusive: "I feel like his game is more and more like the 'Go god'. Really, it is brilliant." *Id.*

Let us assume computer programmers develop a program with artificial intelligence—call it Hercules[15]—that predicts possible legal outcomes more effectively than even the most seasoned and talented attorneys. Would computers then render lawyers obsolete? Not necessarily. There may be more to legal interpretation than just prudence. Another relevant concern may be moral. There may be a *moral* obligation to follow the law—or at least to take it into account in deciding how to act. If there is, how should legal interpreters—presumably lawyers—advise autonomous car manufacturers about what the law requires (or prohibits or permits)? And can Hercules displace them? Could Hercules eliminate any role for lawyers—or other human beings[16]—in making the life and death decisions at issue? Could a computer program the car that will decide whether to mow down pedestrians or sacrifice its driver? Have we arrived at a point where we have law—and legal ethics—without a human mind?

Note the potential parallel to the challenges facing the autonomous car when there is no way to avoid the loss of human life. The car, we noted, must first make a descriptive or predictive assessment of the consequences of different courses of action. Second, the car must make a prescriptive assessment about the right choice is between those options. The same distinction can apply to legal interpretation. A first challenge is describing the law or predicting how others would likely interpret it. Given the indeterminacy in the law, one would expect the result to be various possible interpretations with different likelihoods of being adopted by an authoritative legal interpreter.[17] A second

15. I borrow this name from Ronald Dworkin's fictional idealized interpreter, first introduced in Ronald Dworkin, *Hard Cases*, 88 HARV. L. REV. 1057, 1083 (1975) and later developed in RONALD DWORKIN, LAW'S EMPIRE (1986). Dworkin's Hercules, much like the program in the text, has a capacity to synthesize different sources of legal authority in a way no human being can do.

16. Of course, Hercules would not displace all human beings unless it can make the relevant business decisions. Let's assume it can—at least to the extent those decisions are made to maximize profit and do not entail the sorts of moral judgments that may be necessary for legal interpretation as well. *See, e.g.*, SUSSKIND & SUSSKIND, *supra* note 12, at 78-84 (discussing, inter alia, the ways in which technology has displaced management consultants).

17. There is a parallel here to moral disagreement. Philosophers have been developing "metanormative" theories with the hope of coming up with a mechanism to guide intelligent machines in reconciling conflicting moral theories. *See, e.g.*, Kyle Bogosian, *Implementation of Moral Uncertainty in Intelligent Machines*, 27 MINDS

challenge is selecting a particular legal interpretation from among those available. How *should* the law be interpreted? That may involve some moral judgment—and preserve a role for lawyers and other human beings, unless and until Hercules can make not only descriptive or predictive judgments about the law but also moral ones.[18]

This paper applies a novel jurisprudential thesis to argue that lawyers—and other human beings—would remain relevant even after the rise of Hercules. The thesis is that the best account of the nature of law varies with the purpose of interpretation. More specifically, when legal interpreters seek merely to describe the law—or to predict how others will interpret it—the law is best understood as consistent with legal positivism. However, when legal interpreters look to the law as a source of moral guidance, they must rely on morality to render it sufficiently determinate to be useful.[19] Hence, they must act as non-positivists (or natural lawyers). So if lawyers working with autonomous car manufacturers believe they have a moral obligation to advise their clients to abide by the law—and this paper suggests reasons they might—then they should act as natural lawyers. They should make moral judgments in interpreting the law. And that may justify a continuing, special role for human beings in saying what the law is. Thus, we may never arrive at interpretation of the law—and legal ethics—without a human mind.

Part II focuses on technology. Part II.A first explores AI's potential, providing a brief review of recent developments. It suggests that computers soon may not only drive our cars but also predict how

& MACHINES 591, 595-603 (2017) (discussing, inter alia, William MacAskill, Normative Uncertainty (Feb. 2014) (Ph.D. dissertation, Oxford University), http://commonsenseatheism.com/wp-content/uploads/2014/03/MacAskill-Normative-Uncertainty.pdf). As Bogosian rightly recognizes, disagreements about morality extend to and include disagreements about metamorality. *Id.* at 603-04.

18. If that day arrives, as discussed below, computers may displace lawyers and judges alike.

19. I have begun to develop this thesis elsewhere, at times collaborating with the philosopher Manuel Vargas. *See, e.g.*, Davis, *Legality*, *supra* note 5; Manual Vargas & Joshua P. Davis, *American Legal Realism and Practical Guidance*, *in* REASONS AND INTENTIONS IN LAW AND PRACTICAL AGENCY (G. Pavlakos & V. Rodriguez-Blanco eds., 2015); Joshua P. Davis, *Legal Dualism, Legal Ethics, and Fidelity to Law*, 2016 J. PROF. LAW. 1 (2016); Joshua P. Davis & Manuel R. Vargas, *Legal Dualism, Naturalism, and the Alleged Impossibility of a Theory of Adjudication*, (on file with author).

the law will apply to self-driving vehicles. Part II.B discusses a danger that accompanies the growing role of AI—that computers will make decisions that rely on troubling inferences and have undesirable effects. Part II.C notes a challenge in addressing these issues: the difficulty of determining and understanding how computers make the assessments they do, a seemingly necessary step in avoiding the potential dangers they pose. Part II.D then notes a likely limitation on the role of computers: that they can describe and predict—helping us to choose the means for accomplishing our ends—but they may be unable to make the moral judgments necessary to identify the ultimate ends we should pursue.

Part III then turns to jurisprudence. Part III.A notes an important relationship between the potential role of computers as discussed in Part II.D and jurisprudence: on one hand, computers may not be able to make ultimate value judgments, including moral judgments; on the other hand, the central debate in jurisprudence for the past fifty years has been about the role of moral judgments in saying what the law is.[20] Part III.B reviews a novel solution to this central debate in jurisprudence, one I have been developing in recent years (at times in writings co-authored with the philosopher Manual Vargas): "Legal Dualism."[21] It suggests that morality need not play a role in saying what the law is when a legal interpreter seeks merely to describe the law or to predict how other will interpret it. But that morality does play a necessary role when a legal interpreter seeks moral guidance from the law. Legal Dualism thus identifies a possible limit to the role computers can play in legal interpretation: they cannot replace human beings when the law serves as a source of moral guidance. Part III.C suggests reasons to believe that this point is not purely academic—that law is likely to serve as a source of moral guidance in some tasks, such as programming autonomous cars. Part III.D finally notes that this inquiry into the nexus between technology and jurisprudence suggests a solution to the main controversy in jurisprudence as well a potentially productive new line of inquiry: adopting Legal Dualism may provide a way past the long-running debate between legal positivists and natural lawyers, one that can help us circumscribe the appropriate role for AI in legal interpretation. And redefining the distinction between legal

20. *See* Hershovitz, *supra* note 8, at 1162.

21. *See* Davis, *Legal Dualism*, *supra* note 19.

positivism and non-positivism (or natural law) may advance jurisprudence in a way that tracks the judgments AI can and cannot make.

Part IV provides a brief conclusion. It suggests that contemplation of a world in which computers might serve as lawyers and judges—whether that is a realistic prediction or merely a provocative thought experiment—may teach us something about both AI's nature and the nature of law.

II. COMPUTERS AND LEGAL INTERPRETATION: POTENTIAL, DANGER, CHALLENGE, LIMIT

The growing role computers and AI (or machine learning) play in our society is fascinating and complex. This Article does not attempt to offer any authoritative pronouncements on the subject. Four limited observations, however, are relevant for the present analysis. They address (1) AI's potential, (2) a danger it poses, (3) a challenge that exists in addressing that danger, and (4) an apparent limit to AI's potential. This analysis provides a framework for exploring ways jurisprudence can inform our understanding of AI's proper role in our society and AI can inform our understanding of jurisprudence.

A. Potential: The Inevitability of Computers Interpreting the Law?

Consider first the potential of computers and AI. They tend to outstrip our expectations. Today, they perform analyses that not long ago we thought beyond their reach. Pundits once doubted that a computer would ever beat the World Chess Champion. The human mind—with its intuition—was simply superior. Then it wasn't.[22] Now the top chess players use computers as instructors—as a way to identify new options and to draw inferences about what the best chess moves are.[23] The best computers are superior. Human beings cannot beat them. We can only learn from them.

22. Dana Mackenzie, *Update: Why This Week's Man-Versus-Machine Go Match Doesn't Matter (and What Does)*, SCI. MAG. (Mar. 15, 2016, 10:00 AM), http://www.sciencemag.org/news/2016/03/update-why-week-s-man-versus-machine-go-match-doesn-t-matter-and-what-does (discussing the World Chess Champion's loss to a computer in 1997).

23. *Id.*

Autonomous cars provide another important example. Discussion of self-driving cars is relatively recent—at least among those who do not specialize in technological innovation. The project appeared daunting. The human mind processes and organizes massive amounts of data while operating a motor vehicle. The real world is not a chessboard. Chess pieces operate in a fixed space—an eight by eight grid—with clearly prescribed rules for motion. There are no distracted teenagers texting on cell phones while eating breakfast and switching radio stations. No small children dash unexpectedly in front of a moving pawn or knight. Human cognition—including the ability to perceive and interpret information about the physical world and react in real time—would seem to have a huge advantage over computers. But now many commentators suggest that self-driving cars are safer than have the potential to eliminate "at least 90% of road deaths."[24] They are safer than human drivers, whose displacement may be inevitable. In the not so distant future, it may be difficult to obtain insurance for people who want to drive themselves.[25]

Given this history and context, the prospect of computers interpreting the law—at least in the sense of predicting how courts will rule—does not seem far-fetched.[26] Literary theorists like to tell us that there is no self-interpreting text.[27] Perhaps there isn't. Or perhaps there wasn't. But perhaps there will be. Computers may be able to interpret texts on their own—detecting, amplifying, and clarifying meanings in much the same way human beings do. They may be able to say what texts mean, including legal texts. Or at least they may be able to synthesize human uses of language and predict how human beings

24. TEGMARK, *supra* note 1, at 99.

25. *See id.* at 109 (discussing that particularly safe self-driving cars may be lead to less expensive insurance than for human drivers).

26. *See* SUSSKIND & SUSSKIND, *supra* note 12, at 69-70 ("Big Data techniques are underpinning systems that are better than expert litigators in predicting the results of court decisions, from patent disputes (the Lex Machina service) to the US Supreme Court.") (citing https://lexmachina.com); Daniel M. Katz et al., *Predicting the Behavior of the Supreme Court of the United States: A General Approach*, CORNELL U. LIBR. (July 23, 2014), https://arxiv.org/pdf/1407.6333.pdf.

27. *See, e.g.*, Patrick Sullivan, *"Reception Moments," Modern Literary Theory, and the Teaching of Literature*, 45 J. ADOLESCENT & ADULT LITERACY 568, 568 (2002) ("We now regard the process of creating meaning as a kind of collaboration between the author, the reader, the culture or 'interpretive community' the author and the reader inhabit, and the language with which the text is constructed.").

would interpret and respond to texts, including legal texts.[28] Hercules may rise sooner than we expect.

In this context, consider Franklin Foer's description of Google's aspirations:

> At the epicenter of Google's bulging portfolio is one master project: The company wants to create machines that replicate the human brain, and then advance beyond. This is the essence of its attempts to build an unabridged database of global knowledge and its efforts to train algorithms to become adept at finding patterns, teaching them to discern images and understand language.[29]

According to Foer, Google seeks to recreate the human mind and improve it.

Similarly, Foer believes Facebook aims to displace ordinary government with computers and the engineers that design them.[30] Foer even quotes Mark Zuckerberg, the founder of Facebook, acknowledging, "In a lot of ways Facebook is more like a government than a traditional company. We have this large community of people, and more than other technology companies we're really setting policies."[31]

Google aspires to enable AI to do all the thinking people can do, only more effectively. Facebook may be taking over responsibilities usually reserved for government. And we recently read in the news that AI for the first time has performed better than human beings on a

28. I do not mean to take a position on disagreements within literary theory. I mean only to be suggestive and remain agnostic about what it is that AI would be assessing if and when it can make accurate claims about the meanings of texts, at least for predictive purposes.

29. FOER, *supra* note 1, at 33. If one were to quibble with this summary, one might question whether Google trains algorithms and teaches them, or whether it is more apt to say that the computer programs train and teach themselves.

30. *Id.* at 61. Foer writes of Zuckerberg as inheriting "an abiding fantasy, a dream sequence in which we throw out the bum politicians and replace them with engineers—rule by slide rule." *Id.*

31. *Id.*

reputable test of reading-comprehension test.[32] Computer judges do not seem so far-fetched.[33]

Next note the commentary of Max Tegmark in his recent book, *Life 3.0: Being Human in the Age of Artificial Intelligence:*

> Since the legal process can be abstractly viewed as computation, inputting information about evidence and laws and outputting a decision, some scholars dream of fully automating it with *robojudges:* AI systems that tirelessly apply the same high legal standards to every judgment without succumbing to human errors such as bias, fatigue or lack of the latest knowledge.[34]

Tegmark implies that the judge's role is purely mechanical, if complex. An ideal judge engages merely in computation. If so, AI judges—what he calls robojudges—might have several advantages. In theory, they would not be affected by subconscious bias.[35] Moreover, they could be replicated and indefatigable, so we would not need to worry about having too few judges or exhausting the ones we have.[36] And they hold the potential for unlimited memory and learning capacity, so we would not be concerned with lack of expertise or knowledge.[37] Enter Hercules (or many Herculeses).

To be fair, Tegmark also recognizes potential liabilities of robojudges. They might get hacked.[38] They also might lack transparency, undermining respect for the legal system.[39] And they might not cure but rather replicate patterns of discrimination, as

32. Sherisse Pham, *Computers Are Getting Better than Humans at Reading*, CNN: TECH (Jan. 16, 2018, 4:16 AM), http://money.cnn.com/2018/01/15/technology/reading-robot-alibaba-microsoft-stanford/index.html.

33. *See, e.g.*, Anna Ronkainen, *From Spelling Checkers to Robot Judges?: Some Implications of Normativity in Language Technology and AI & Law*, SSRN (July 6, 2011), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1879426&download=yes (discussing the possible use of software in judging and practical opportunities and challenges).

34. TEGMARK, *supra* note 1, at 105.

35. *Id.* at 105.

36. *Id.* at 106.

37. *Id.*

38. *Id.*

39. *Id.*

evidenced by a recent study that showed software designed to predict recidivism resulted in bias against African Americans in sentencing.[40]

Tegmark's vision is as awe-inspiring as it is frightening. It makes us wonder whether there is anything human beings can do with their minds that computers will not soon be able to do better. It also should cause us to contemplate the risks to which computer "thinking" may give rise.

B. Danger: Do Computers Act for Improper Reasons?

The expanding role of computers and AI in decision-making should fill us with awe, in both its positive sense—awesome—and negative sense—awful. Consider how one commentator—Frank Lautz, the director of NYU's Game Center—reacted to a new chess program, AlphaZero, developed by DeepMind, a secretive artificial intelligence subsidiary of Google.[41] AlphaZero may well be the best chess player in the world. Moreover, it is striking in another way. It does not borrow from centuries of human experience playing chess—as did many of its predecessors, including GO AI—but rather builds its algorithms from scratch.[42] Mr. Lautz's response:

> For a while, for like two months, we could say to ourselves, "Well, the Go AI contains thousands of years of accumulated human thinking, all the rolled up knowledge of heuristics and proverbs and famous games." We can't tell that story anymore. If you don't find this terrifying, at least a little, you are made of stronger stuff than me. I find it terrifying, but I also find it beautiful. Everything surprising is beautiful in a way.[43]

Mr. Lautz's terror may be more understandable in other contexts. After all, chess is just a game. But we are likely all familiar with the

40. *Id.* at 106-07 n.36 (citing Julia Angwin et al., *Machine Bias*, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing; O'NEIL, *supra* note 1, at 204 (noting how "Big Data" perpetuates inequality by "codify[ing] the past" rather than "invent[ing] the future").

41. *See* Oliver Roeder, *Chess's New Best Player Is a Fearless, Swashbuckling Algorithm*, FIVETHIRTYEIGHT (Jan. 3, 2018, 12:52 PM), https://fivethirtyeight.com/features/chesss-new-best-player-is-a-fearless-swashbuckling-algorithm/.

42. *Id.*

43. *Id.*

role that computer algorithms may have played in subverting the presidential election of 2016. Russians may have "hacked" social networks, including Facebook and Twitter, to alter the political preferences of many Americans.[44] That is not the only danger. Consider Harvard Professor Latanya Sweeney's study showing that "African American names were frequently targeted with Google ads that bluntly suggested that they had arrest records in need of expunging."[45]

A problem is that computers do not rule out forms of analysis that may be immoral, unethical, or illegal. Imagine, for example, a firm—call it AllTooCommon Corp. ("ATCC")—that has a troubling pattern of sexual harassment. Assume women at ATTC regularly experience inappropriate behavior and their performance suffers. Some of them say nothing and become disaffected. Others report the behavior and suffer retaliation. Either way, assume women disproportionately experience illegal, adverse employment decisions. They perform worse than men by apparently objective measures because of unlawful conduct directed at them in the workplace.[46]

Enter a computer charged with predicting performance for purposes of hiring, retention, compensation, and promotion. It uses AI. It does not simply apply an algorithm. It detects patterns and uses them to generate and adapt new algorithms. Given these circumstances, the computer may well reinforce the discrimination already occurring in the workplace. Using the information available about employees and whatever metrics the company uses for measuring performance—likely including past evaluations, retention, raises, and promotions—the computer would be apt to predict that, all else equal, women will

44. Scott Shane, *The Fake Americans Russia Created to Influence the Election*, N.Y. TIMES (Sept. 7, 2017), https://www.nytimes.com/2017/09/07/us/politics/russia-facebook-twitter-election.html.

45. FOER, *supra* note 1, at 71.

46. Given the recent string of stories in the news, this hypothetical seems painfully plausible. As one example among many, consider the alleged culture of harassment at two Chicago Ford plants as reported by the New York Times. Susan Chira & Catrin Einhorn, *How Tough Is It to Change a Culture of Harassment? Ask Women at Ford*, N.Y. TIMES (Dec. 19, 2017), https://www.nytimes.com/interactive/2017/12/19/us/ford-chicago-sexual-harassment.html. Ford provides one among too many examples. *See* Emily Steel, *At Vice, Cutting-Edge Media and Allegations of Old-School Sexual Harassment*, N.Y. TIMES (Dec. 23, 2017), https://www.nytimes.com/2017/12/23/business/media/vice-sexual-harassment.html.

perform less well than men. By so doing the computer would in effect punish women for the discrimination they have suffered, seemingly discriminating impermissibly on the basis of sex.[47]

This example is no mere conjecture. Companies are already using AI to assess their employees.[48] And commentators are already expressing concern that AI may thereby reinforce biases—and may do so in ways that make legal redress difficult.[49] Our current legal doctrines do not necessarily lend themselves to policing companies that rely on AI, even when the AI relies on analyses that might well be impermissible if undertaken by human beings. Is there ever discriminatory intent, for example, when it comes to data mining and artificial reasoning? And are such efforts—by design—necessarily job-related and consistent with business necessity?[50]

Similar risks may beset the decisions of self-driving cars, if less obviously so. Consider what would happen if a car were programmed to minimize the legal liability it causes from an accident.[51] Of course, making that assessment in real time would be a daunting, technical task. But assume a car's computer can do it. It can appraise the value of other vehicles on the road, the likelihood of harm to drivers, passengers, and pedestrians, and the resulting potential liability, including from lost income. Given the allocation of capital and earnings in our society, it would not be difficult to predict whose lives the car would value more and whose it would value less. All else equal, drivers of more expensive vehicles would fare better than drivers of less expensive vehicles. The

47. Similar potential and danger accompany other uses of computer algorithms and artificial intelligence. Consider the reliance of Pittsburgh officials at child protective services on computers to predict which children are in danger. *See* Dan Hurley, *Can an Algorithm Tell When Kids Are in Danger?*, N.Y. TIMES MAG. (Jan. 2, 2018), https://www.nytimes.com/2018/01/02/magazine/can-an-algorithm-tell-when-kids-are-in-danger.html. Protecting children from domestic violence is about as compelling a reason as one can find to rely on technology. That said, it was only a matter of time before the decisions made in Pittsburgh—or elsewhere—gave rise to perceived patterns of improper discrimination, including along the lines of race. *See* EUBANKS, *supra* note 1; Weisberg, *supra* note 4, at 45-47 (reviewing EUBANKS, *supra* note 1).

48. *See* Kim, *supra* note 3, at 902-09.

49. *Id.*

50. *Id.*

51. *See* Casey, *supra* note 9, at 1350 (arguing that profit-maximizing firms will design AI to minimize legal liability, including in the design of self-driving cars).

wealthy would be protected. And the wealthy are likely to be white. Further—again, all else equal—people in their prime years of earning would be treated more favorably than the young and the old, men more favorably than women, and white people more favorably than people of color. The car would make life and death decisions in a way that reinforces inequalities along the lines of class, age, sex, and race.[52]

The same might be true for legal interpretation. Indeed, the law creates the incentives that could cause a self-driving car to act in potentially objectionable ways. Related and more general concerns might arise from a computer interpreting the law, perhaps by predicting how judges would interpret it. Judges are disproportionately white men. Do they have biases that affect the patterns of their decision-making? Do judges from other groups also have biases, perhaps running in different directions? Do those patterns predict judicial decisions, perhaps even in ways the judges may not recognize, and human legal interpreters might not notice or take into account? If so, Hercules is likely to detect those trends—to recognize patterns of discrimination in case law—and may well offer legal interpretations that reinforce them. Hercules might even do so to a greater extent than would practicing lawyers, who might seek to cleanse the law of its impurities.[53] Maybe Mr. Lautz is right—we should all be at least a little terrified.

C. Challenge: Is Artificial Intelligence Ineluctably Inscrutable?

More daunting yet is that we may have a limited ability to perceive, much less correct, inequities arising from AI. One reason is that AI often is not transparent. If it were, we might be able to cull undesirable biases from computer decision-making. If a computer were able to indicate that it weighed a woman's gender to a specified extent against her in predicting her future job performance, it might be possible to

52. *See, e.g.*, MARTHA CHAMALLAS & JENNIFER B. WRIGGINS, THE MEASURE OF INJURY: RACE, GENDER, AND TORT LAW (2010). Note that Judge Jack Weinstein has issued rulings that resist the general trend in tort law, preventing lawyers from arguing for reduced liability based on race or ethnicity. Ashley Southall, *Award in Lead Paint Lawsuit Can't Be Tied to Ethnicity, Judge Rules*, N.Y. TIMES (July 29, 2015), https://www.nytimes.com/2015/07/30/nyregion/award-in-lead-paint-lawsuit-cant-be-tied-to-ethnicity-judge-rules.html.

53. *See* Davis & Vargas, *Legal Dualism, Naturalism, and the Alleged Impossibility of a Theory of Adjudication*, *supra* note 19 (forthcoming 2019).

excise that strand of the analysis—to correct for it. That, however, may not be possible in practice for several reasons.

One difficulty in ameliorating how computers reason is that there may be insufficient time. We have already put an extraordinary burden on self-driving cars. For example, we require them to operate safely without human oversight and, in moments of crisis, to minimize harm using some workable set of criteria. It may not be possible to superimpose yet another layer of analysis—requiring cars to assess whether the safety algorithms they keep developing and adapting contain implicit biases. If that process were to involve human supervision, it would not be possible in real time. Indeed, that would seem to defeat the purpose of making cars autonomous. On the other hand, if the cars were to monitor themselves, yet another layer of AI would be necessary. That might not prove feasible. And the additional layer of AI would presumably require oversight too.

Another reason for the opacity of AI is that it creates and adapts its own algorithms on an ongoing basis, so that even those who design it are not in the ordinary course able to predict how it will do what it does or to assess after the fact what it did. Consider the plight of Michal Kosinski, a professor at the Stanford Graduate School of Business. He applied an open-source facial-recognition algorithm to publicly posted dating profiles.[54] He was interested in exploring a fraught subject—whether physical characteristics correlate with personality traits.[55] At first he found nothing interesting. However, when he asked the algorithm to use photographs to identify the sexual orientation of subjects, it did so with 91 percent accuracy for men and 83 percent accuracy for women.[56] That result was both startling and disturbing. When he published his study, it gave rise to a predictable outcry.[57] What else might artificial intelligence detect from photographs? What might that mean for privacy in the future?

What received less attention—as reported by Cliff Kuang for The New York Times Magazine—was "a genuine mystery that went almost

54. Cliff Huang, *Can A.I. Be Taught to Explain Itself?*, N.Y. TIMES MAG. (Nov. 21, 2017), https://www.nytimes.com/2017/11/21/magazine/can-ai-be-taught-to-explain-itself.html.

55. *See id.*

56. *Id.*

57. *Id.*

ignored amid all the media response: *How* was the computer doing what it did? What was it seeing that humans could not?"[58] Even Kosinski did not know. The algorithm was not designed to reveal the patterns it detected. Kosinski was left to conduct various experiments to *infer* as best he could how the algorithm did what it did.[59]

Finally, and most speculative, the way a computer "thinks" may not be susceptible to human understanding. It is at least possible that a computer would organize information in a manner that human beings cannot understand—perhaps because we are not smart enough, perhaps because certain patterns are not meaningful to us, perhaps because there are certain things we cannot perceive, or perhaps for other reasons.

This last point has the potential to be profound and slippery, requiring careful work in epistemology and ontology. However, this Article is not the place to undertake that effort. For present purposes, it suffices to recognize that time and transparency may not be enough for human beings to monitor AI. Something more may be required. That point may seem inaccessibly abstract. But it isn't necessarily so. Try to explain probability, calculus, or multiple regression analysis to most eight-year-olds. They simply cannot understand, regardless of whether the explanation is clear and complete and the theory sound. The concepts are just too difficult. The same is true for quantum mechanics and relativity—for many adults, not just children. And the same may be true for concepts that AI may develop and use. Perhaps human beings—even the smartest and most knowledgeable among us—will be incapable of understanding them.

The three points above suggest some of the difficulties that may arise under the European Union's General Data Protection Regulation ("GDPR").[60] The GDPR includes a "right to explanation"—the right to demand an explanation for how an algorithm reached its conclusions.[61] What this will mean in practice is unclear. Government

58. *Id.* (emphasis in original)

59. *Id.*; *see also* Weisberg, *supra* note 4 (discussing Kosinski's work as well as a similar mystery about how a German "handwriting recognition algorithm can predict with 80 percent accuracy whether a sample was penned by a man or woman").

60. *See* EU GDPR, https://eugdpr.org/ (last visited Dec. 3, 2018). The regulation took effect on May 25, 2018. *Id.*

61. J.M. Porup, *What Does the GDPR and the "Right to Explanation" Mean for AI?*, CSO (Feb. 9, 2018, 3:16 AM), https://www.csoonline.com/article/3254130/compliance/what-does-the-gdpr-and-the-right-to-explanation-mean-for-

officials will have a difficult time defining what counts as a sufficient explanation. Does it have to be understandable? If so, by whom? The average person on the street? An expert? A few specialists in the relevant area? A hypothetical person with sufficient knowledge and intelligence to understand the explanation, even if no one in fact possesses either? Does it just have to be technically correct and complete, even if abstruse—beyond human reckoning?

In sum, it is hard to know how practical the "right to explanation" will be. Some decisions—like those made by self-driving cars—may be swift and irrevocable. Piecing together the computer's "thinking" after the fact may have limited utility. Moreover, the task of developing programs that could reveal how the computer made its decision may not be feasible. That undertaking may be so expensive and cumbersome that it could come at too high a cost to innovation. And, in any case, human beings may not be capable of understanding the computer's "reasoning," even if the computer is in some sense capable of providing an explanation. This last point in particular may apply to legal interpretation. Hercules may in some ways be able to interpret the law more effectively than human beings, but it may not be able to enlighten us about how it reached its conclusions.[62]

---

ai.html. Recital 71, which accompanies the GDPR and is not legally enforceable, mentions a "right to explanation." *Id.* And the GDPR itself states that data controllers must notify consumers how their data will be used, including "the existence of automated decision-making" and, at least in some circumstances, "meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject." EU GDPR, *Article 15: Right of Access by the Data Subject*, PRIVAZYPLAN, http://www.privacy-regulation .eu/en/article-15-right-of-access-by-the-data-subject-GDPR.htm (last visited Dec. 3, 2018).

62. This may be true even if Hercules can justify its decisions in language that human legal interpreters use. Note the distinction—often drawn by Brian Leiter, for example—between the actual motivation for a legal interpretation, including a judicial decision, and the justification a legal interpreter, such as a judge, offers for it. *See, e.g.*, Brian Leiter, *Explaining Theoretical Disagreement*, 76 U. CHI. L. REV. 1215, 1224-25 (2009) (explaining the "Disingenuity Theory," that legal interpreters, including judges, may be disingenuous about the basis for reaching their conclusions, in particular that they may claim to be following the law when in fact they are creating new law).

D. Limit: Do Autonomous Cars Lack Autonomy?

So the potential for AI may be awesome, but it also may be terrifying in ways that are not easily addressed. That does not mean their potential is limitless. In particular, computers and AI so far have been able to address only means, not ultimate ends. Put differently, they have operated in the realm of fact, not value.[63] They must be provided ultimate objectives.[64] Once they are, they can be used to detect patterns and predict behavior, often far more effectively than human beings. They may even be able to adopt intermediate or instrumental goals—ones that can help them achieve the ultimate goals that they have been assigned. As of yet, however, they cannot select the goals they should attempt to achieve. In a sense the term "autonomous car" may thus be a misnomer. We often conceive of autonomy as involving a choice of ends, not just of means. We have not yet reached the point where computers choose ends. And it is not clear that we will.[65]

The prospect of empowering computers to make moral judgments is not particularly promising. After all, there is little agreement in moral philosophy about the right way to frame moral problems. Philosophers dispute the nature of moral propositions, how they can or should be tested, how people gain moral insights, whether moral claims can be true or false, what the organizing principles of morality are, and any number of related issues. It is ironic, no doubt, that the very limited ability of human beings to make progress toward consensus in moral philosophy and to make confident moral judgments may render them superior to computers in both regards. We may not be able to do better than to muddle our way through these murky waters on our own—without technological assistance—in part because we do not have

63. *See*, *e.g.*, RONALD DWORKIN, RELIGION WITHOUT GOD (2013) (providing a general discussion of this distinction).

64. I say "ultimate" objectives because AI would presumably be able to identify instrumental objectives that assist in achieving a specified set of ultimate ones. *See* TEGMARK, *supra* note 1, at 264 (discussing the relationship of subgoals and ultimate goals); *see also* BOSTROM, *supra* note 1, at 132 (discussing the instrumental convergence thesis).

65. For example, Max Tegmark suggests four principles that he hopes distills human morality to a workable core: utilitarianism; diversity; autonomy; and legacy. TEGMARK, *supra* note 1, at 271.

sufficient understanding to provide computers clear enough directions to help us.[66]

This article does not seek to resolve the knotty issue of whether computer can or will be able to render moral judgments. But a preliminary sketch of some of the challenges seem in order. Three basic options are available: (1) a top-down approach,[67] (2) a bottom-up approach,[68] and (3) a predictive approach. The top-down approach would involve providing a general principle or a set of general principles to guide AI decisions.[69] One problem is that there is nothing close to a consensus—among philosophers, ordinary citizens, elected officials, or seemingly any other relevant group—about what the right moral principles are.[70] Moreover, in trying to choose among the relevant moral principles in a particular case, there is likely to be a need for what one might call "local" moral judgments—judgments that are sensitive to context or setting. Purely abstract moral judgments are likely to prove insufficient to guide human conduct or AI.[71] The need for local moral judgments renders a top-down approach to morality unlikely to succeed.

---

66. *See, e.g.*, Bogosian, *supra* note 17, at 595-603 (discussing the efforts to overcome this challenge).

67. *See* WENDELL WALLACH & COLIN ALLEN, MORAL MACHINES: TEACHING ROBOTS RIGHT FROM WRONG 83-97 (2009) (discussing the possibilities—and challenges—regarding a top-down system).

68. *See id.* at 99-115 (discussing the possibilities—and challenges—regarding a bottom-up approach). Wallach and Allen also discuss a hybrid of the two. *Id.* at 117-24.

69. *Id.* at 83.

70. *See, e.g.*, Bogosian, *supra* note 17, at 592 (noting in 2009 survey 26% of philosophy faculty members accepted or leaned toward deontology, 24% consequentialism, and 18% virtue ethics, with the remaining 32% favoring other approaches). Of course, disagreement about morality poses a similar problem for people. We rely in part on our democratic processes to overcome this problem—relying on elected representatives to make law and to appoint judges to render appropriate judgments in interpreting and applying the law, or we elect judges to make those decisions.

71. For this reason, many philosophers have endorsed the notion of a reflective equilibrium—an iterative process of informing principles or rules with judgments about particular cases and vice versa. *See, e.g.*, JOHN RAWLS, A THEORY OF JUSTICE (1971). The literature on this topic is large. *See, e.g.*, NORMAN DANIELS, JUSTICE AND JUSTIFICATION: REFLECTIVE EQUILIBRIUM IN THEORY AND PRACTICE (1996).

The second approach—bottom-up—faces similar challenges. A bottom-up approach would involve induction rather than deduction—with AI developing its own moral commitments through experience.[72] AI would not be programmed to incorporate general moral principles, but rather would discern them—or, at least, recognize moral actions or outcomes—in specific contexts, perhaps undergoing a process akin to human moral development.[73] For that approach to work, AI would have to receive input about the right moral outcome or action in particular cases or would have to develop some capacity to make the relevant moral judgments itself. This second approach would give rise to myriad problems. One involves the intertwinement of moral (or value) judgments and other reasons for action, including prudence, heuristics, and biases. It is unclear even in principle how a computer could disentangle the various reasons to endorse or reject a particular result or course of conduct. Further, if a computer were to develop its own moral judgment, it might need to acquire subjective experience or something very close to it—so as to feel necessary emotions and empathy.[74] But that subjective or quasi-subjective experience might lead to moral and other judgments that we find unacceptable—perhaps a technocentric worldview that does not comport with our anthropocentricism.[75]

The third approach—prediction—would involve a computer moral or other value rather than making substantive moral or other value judgments itself. That approach is reminiscent of Holmes' Bad Man theory and his famous proclamation, "The prophecies of what courts will do in fact, and nothing more pretentious, are what I mean by the

72. WALLACH & ALLEN, *supra* note 67, at 99-101.

73. *Id.*

74. There exists a range of views on the potential role of emotions in moral reasoning. *See, e.g.*, PAUL BLOOM, AGAINST EMPATHY: THE CASE FOR RATIONAL COMPASSION (2016); JEFFREY GREENE, MORAL TRIBES: EMOTION, REASON, AND THE GAP BETWEEN US AND THEM (2013); MORALITY AND THE EMOTIONS (Carla Bagnoli ed., 2011); MARTHA C. NUSSBAUM, UPHEAVALS OF THOUGHT: THE INTELLIGENCE OF EMOTIONS (2003).

75. In fairness, our anthropocentricism may lack a sound moral basis. *See, e.g.*, PETER SINGER, ANIMAL LIBERATION (1975); Peter Singer, *Equality for Animals?*, *in* PRACTICAL ETHICS 48 (1979).

law."[76] Paraphrasing Holmes, we might say the prophecies of the moral judgments people will make in fact, and nothing more pretentious, is what AI means by morality. Given the parallel to Holmes' Bad Man theory, the predictive approach may be subject to the powerful critique of that theory leveled by H.L.A. Hart in "The Concept of Law." As Hart argued—among other points—a judge would perform her job poorly if she sought to predict how she herself would rule in a case, an inquiry that seems hopelessly circular.[77] But note that AI is not being asked simply to predict how a court will rule. It is being asked to predict the *moral* judgments judges—or other legal interpreters—would make (to the extent necessary to engage in legal interpretation). That provides at least a partial response to Hart that was unavailable to Holmes. Still, the two main points discussed above pertain to the third approach as well. The local nature of moral judgments and the way in which moral reasons are inextricably intertwined with other reasons for action pose formidable obstacles to AI predicting and mimicking the moral judgments that would be expected from human interpreters.

In exploring the predictive approach, it may prove helpful to distinguish the perspectives of judges and attorneys. Consider first a robojudge. Because moral judgments are local, the robojudge would likely need a significant stock of judicial rulings—of data points reflecting decisions closely related to the one before the robojudge—to mimic a human judge. General moral principles gleaned from ordinary human life or other institutional settings would not likely suffice. Those other contexts would provide only limited guidance, as the relationship between them and the judicial context would be attenuated. So robojudges would not seem capable of displacing human judges entirely. If they did, the data on which they rely would be grow less and less pertinent over time to the cases that come before the robojudges and the quality of their moral judgments would degrade.

Further, a robojudge could have difficulty separating the moral judgments immanent in judicial opinions from the other reasons—

76. Oliver W. Holmes, Jr., *The Path of the Law*, 10 HARV. L. REV. 457, 459 (1897).

77. H.L.A. HART, THE CONCEPT OF LAW 124-54 (1961).

Further, given the indeterminacy of the law, a prediction theory will result in a range of outcomes that might occur and a likelihood of each one. It is not clear how AI should choose between them. As discussed below, the most likely outcome is not necessarily the most attractive one for various reasons.

prudence, heuristics, biases, *etc*.—that motivate them. To the extent a robojudge should predict moral judgments—and not just predict how courts will rule—it is important for AI to distill only the role of morality—and ignore other reasons or motivations—from the pattern of judicial decision-making. But it is not clear how AI could go about that task. Although AI may become proficient at identifying patterns in human decision-making, it is not clear how it could distinguish the moral bases of human reasoning from other bases without making independent moral judgments—the very task the prediction approach is meant to avoid.

The above analysis suggests that the predictive approach provides only a second-best approximation of the moral judgments necessary for legal interpretation by a judge. It relies on human decisions for the data necessary to guide a robojudge. That data about moral judgments will be intertwined with other, non-moral influences on judicial decision-making. Both of these phenomena would likely cause the predictions of robojudges to deviate from the moral judgments human beings would make.[78]

Now consider robolawyers. Replacing human lawyers with AI by itself would not create a risk of depriving robolawyers of the data needed to detect patterns in judicial decision-making. As long as people serve as judges, that data would remain available. But a different problem arises. Judges operate within institutional structures that are different than the ones in which lawyers operate. The legal conclusion that a judge reaches is not necessarily the same one a lawyer should reach. A judge, for example, might dismiss a criminal case for lack of evidence. But, as critics of Holmes' Bad Man theory note, it seems wrong to say that it is legal to commit murder as long as a one leaves behind insufficient evidence to support a conviction. And it would be unethical for an attorney to assist a client in committing a murder in

78. I acknowledge the thorny problem of whether human beings would make better moral judgments than AI. One could even imagine, given disagreements over moral judgments, that random variations by AI from human judgments might be as likely to approximate correct moral judgments—if there are correct moral judgments—more closely rather than less closely. On other hand, such a skeptical view would seem to encourage us to give up on attempting to make moral judgments entirely.

such a way as to leave no evidentiary trail and get away with it.[79] So the outcomes in judicial proceedings do not necessarily provide lawyers direct guidance , for example, about the advice they may give their clients or the conduct they may ethically assist. That gives rise to a problem in light of the local nature of moral judgments. The judicial context may be too distant from the lawyerly context to support inferences about moral judgments that translate from one to the other. Meanwhile, the same difficulties that beset robojudges in disentangling moral judgments from other judgments would apply also to a robolawyer.[80]

In any case, a computer's capacity to make moral judgments has lagged well behind its technical ability to describe and predict patterns of behavior. AI has defeated the greatest human genius at Go, but it has

79. *See, e.g.*, MODEL RULES OF PROF'L CONDUCT r. 1.2(d) (AM. BAR ASS'N 1983) ("A lawyer shall not counsel a client to engage, or assist a client, in conduct that the lawyer knows is criminal or fraudulent . . . ."). Bryan Casey appears to argue that the right approach to the issue of morality and AI is to adopt the perspective of Holmes' Bad Man. Casey, *supra* note 9, at 1361-65. To avoid remaining "hopelessly mired," he implies, we should stop trying to make moral judgments about what AI should do. *Id.* at 1347. It is unclear whether his claim is positive or normative. If he is merely predicting how AI will evolve, he is likely right that financial incentives—including those created by the law—will have a big role to play. But if his point is prescriptive—if his argument is that AI companies *should* seek only to minimize their legal liability—his position is subject to the many powerful criticisms that have been leveled against Holmes' Bad Man theory and the crude form of legal positivism it implies. Prominent among them is the rejection of the notion, discussed in the text, that a person who gets away with murder has still violated the law (and morality), properly understood.

80. The analysis in the text focuses on whether AI can make accurate moral judgments. There is another potential reason why computers may not be able to serve in place of human beings as judges or lawyers, at least under some circumstances. This alternative reason would focus not on *what* conclusions a legal interpreter reaches but on *who* the legal interpreter is. This distinction roughly tracks the one Paul Kahn offered between reason and will. *See* Paul Kahn, *Reason and Will in the Origins of American Constitutionalism*, 98 YALE L.J. 449 (1989). We might not let AIs serve as judges for much the reason we might not let them vote—not (just) because of the quality of the decisions they might make but because they are not the sorts of beings that should be given that kind of role in a democratic society. *See* JERRY KAPLAN, ARTIFICIAL INTELLIGENCE: WHAT EVERYONE NEEDS TO KNOW 98-101 (2016) (discussing a thought experiment in which a citizen delegates progressively more power to AI to vote on his behalf). I am grateful to Bradley Wendel for a discussion that suggested this point, one we are exploring together in a separate paper. Developing this point is beyond the scope of this article.

not learned to make the kinds of cogent arguments that would persuade a moral philosopher. Consider in this regard the chapter in Tegmark's book, *Life 3.0*, on "Goals."[81] He provides a perceptive and thought-provoking analysis of various challenges in aligning human goals with AI, breaking that task into three components: (1) *teaching* AI our goals; (2) getting AI to *adopt* our goals; and (3) ensuring AI *retains* our goals.[82] In this discussion, he suggests ways in which AI can learn on its own. He notes, for example, that an eldercare robot might be able to infer from a retired man's activities what he values.[83] When it comes to ethics, however, Tegmark does not suggest a role for AI's self-learning. Note in this regard that merely observing behavior would have limited utility. Most of us do not always act morally. The retired man may be a hypocrite, a liar, a cheat, and a thief. If so, the robot would presumably learn to help the man in his hypocrisy, lying, cheating, and theft. That may be so even if the man acknowledges that those activities are morally wrong.

So how, then, would Tegmark identify the ethical aims AI should pursue? His answer, at least implicitly, is to rely on human beings. As he explains, "postponing work on ethical issues until after goal-aligned superintelligence is built would be irresponsible and potentially disastrous."[84] But he does not suggest ways in which *AI* might identify the content of morality for us. Rather he suggests four principles that *he* distills from his readings over the years: (1) utilitarianism; (2) diversity; (3) autonomy; and (4) legacy.[85] Putting aside the merits of these

81. TEGMARK, *supra* note 1, at 249-80.

82. *Id.* at 260.

83. *Id.* at 261.

84. *Id.* at 269.

85. *Id.* at 271. Tegmark's summary of the principles:

> • Utilitarianism: positive conscious experiences should be maximized and suffering should be minimized.
> • Diversity: a diverse set of positive experiences is better than many repetitions of the same experience, even if the latter has been identified as the most positive experience possible.
> • Autonomy: conscious entities/societies should have the freedom to pursue their own goals unless this conflicts with an overriding principle.
> • Legacy: compatibility with scenarios that most humans *today* would view as happy, incompatibility with scenarios that essentially all humans *today* would view as terrible.

Id.

principles—each taken in isolation and the combination would surely be controversial among moral philosophers, government officials, and ordinary citizens[86]—what is striking for present purposes is that in a book about AI and machine learning, Tegmark has not suggested that we give AI the task of resolving the core ethical issues AI itself raises. It seems we must do that for ourselves.[87]

AI's present and perhaps future inability to make the substantive moral judgments necessary for identifying ultimate ends has implications for the various issues we have discussed so far. Consider sex discrimination in the workplace. As noted, a computer program tasked with predicting success in the workplace might well reinforce and perpetuate sexism. But the fault may lie not with the computer—or at least not only with the computer. Rather human beings are at least partially culpable—assuming there is culpability—for relying on tainted data and for specifying the goals for computers to pursue with insufficient completeness. AI cannot at present determine on its own that one ultimate end is to avoid—and correct for—invidious discrimination.

The same is true for those who provide the criteria that guide driverless cars. Asking a car to minimize legal liability embeds questionable value judgments in its evolving algorithm. The predictable result may well be to perpetuate patterns of discrimination in our society. What is particularly tricky is that goals that seem

86. Obvious points of contention would include Tegmark's decision to value all conscious experiences, including those of AI and not just of organic life forms, and his failure to indicate how positive conscious experiences should be weighed against one another, which could vary based on, among other things, the kinds of experiences (are all pleasures equally valuable, the perverse, the profane, and the pure?) and the entities experiencing them (is the pleasure or pain of a person equivalent to that of a fish or a computer program, assuming AI develops subjective experience?). Tegmark's four principles also seem to embody both utilitarian and deontological commitments, which invites criticism from utilitarians and deontologists alike. The literature exploring, supporting and criticizing each of these moral philosophies is vast and suggest countless questions and concerns about Tegmark's approach. An analysis of those questions and concerns is beyond the scope of this paper.

87. Another way to put the point is that we are left to our own devices—but that seems a confusing metaphor in the circumstances.

Tegmark does acknowledge the possibility that AI will develop consciousness and acquire the full set of rights and responsibilities that human beings have. *Id.* at 276, 281-315.

innocuous—that at first blush appear noncontroversial—may prove questionable or even unacceptable in practice.

We have already questioned the wisdom of asking a self-driving car to minimize legal liability. Consider what may be a more attractive alternative: minimizing expected loss of life. Such an approach would treat all lives equally, unlike the incentives created by our current tort system. But do we really want to do that? Imagine a self-driving car detects that a crash is inevitable because a drunk driver has veered into oncoming traffic. The car faces two options: (1) *almost certainly* kill the drunk driver and *probably* save the life of an innocent child crossing the street; or (2) *almost certainly* save the life of the drunk driver and *almost certainly* kill the innocent child. If the car is minimizing the expected loss of life, it would presumably spare the drunk driver and kill the innocent child. After all, that strategy on average would preserve the most lives.[88] Yet given the culpability of the drunk driver—and the innocence of the child—that result might not be correct.

A similar analysis can apply to a computer engaging in legal interpretation. As noted above, if there is a pattern of bias in judicial decision-making, the computer may embed that pattern—would be expected to embed that pattern—in the predictions it makes. Consider how a computer should address the pervasive indeterminacy that most modern legal scholars believe exists in the law.[89] Mere description or prediction will not—under circumstances of indeterminacy—yield a single interpretation. It will produce numerous interpretive options,

88. To make this example a bit more concrete—if falsely precise—assume in (1) the chance of the drunk driver surviving is 5% and the chance of the child surviving is 55% and in (2) the chance of the drunk driver surviving is 95% and the chance of the child surviving is 5%. Choice (1) would save 0.6 lives on average and choice (2) would save 1.0 lives on average. From the perspective of minimizing the expected loss of life—treating all lives equally—(2) is superior to (1). This hypothetical could be complicated if one takes into account life expectancy, and the goal is interpreted as maximizing expected remaining years of human life. But that doesn't change the main point: the life of a drunk and culpable driver is treated with equal value as the life of an innocent person.

89. Arguably it is in this sense that we are "all realists now." *See* Michael S. Green, *Legal Realism as Theory of Law*, 46 WM. & MARY L. REV. 1915, 1917 (2005) ("[I]t is often said—indeed so often said that has become a cliché to call it a 'cliché'—that we are all realists now."). Most scholars disagree more about the degree of uncertainty in the law rather than its existence.

perhaps with various odds of being selected by a legal interpreter. How is a computer to choose among them?

One possibility might be for the computer to choose the most likely or common legal interpretation. That has a facial appeal. But whether it is the best approach itself requires a moral judgment. It is not obvious that the most popular choice will tend to be the "right" one—depending on our definition of "right"—particularly if we worry that popularity may result from suspect causes. Selecting the most popular outcome may seem unobjectionable when stated abstractly, but it may lead to objectionable results in practice. Perhaps judges or other legal interpreters are prone to make systematic errors, to suffer from common misunderstandings, to fall prey to predictable cognitive biases, or even to act on common undesirable, unconscious prejudices. The popular view of the law—or the view most likely to be popular—may in an important sense be wrong and predictably so. One of the reasons we insulate our federal judges from direct electoral accountability may be based on the view that what is popular is not necessarily what is right when it comes to the law.[90]

A competing approach might look something like the one famously championed by Ronald Dworkin. He claimed that interpretation of the law entails two judgments, one he called "fit" and the other he called "justification."[91] According to Dworkin, interpretation involves assessing how well a plausible result "fits" the law—which one might interpret as being compatible with a computer's prediction about the behavior of judges.[92] But interpretation also involves evaluating how well a result "justifies" the law—that is, how morally attractive it renders the law.

---

90. This position may not be elitist in the way it at first seems. Judges may be no better at making moral judgments than ordinary citizens. But they may be sensitive to institutional concerns that would not be apparent to others without legal training and experience. Further, judges may focus on the broader principles at issue in particular situations rather than on the circumstances of a particular case, applying norms that would—and do—have broad and enduring support in the populace at large. *See* RONALD DWORKIN, *The Forum of Principle*, *in* A MATTER OF PRINCIPLE 33 (1985); CHRISTOPHER EISGRUBER, CONSTITUTIONAL SELF-GOVERNMENT (2001).

91. *See* Davis, *Legality*, *supra* note 5, at 94-95 (providing a brief summary of Dworkin's approach); Joshua P. Davis, *Cardozo's Judicial Craft and What Cases Come to Mean*, 68 N.Y.U. L. REV. 777, 809-10 (1993) (citing DWORKIN, *supra* note 15, at 245-47).

92. *Id.*

Dworkin's approach would have legal interpreters balance fit and justification.[93] That is the task Dworkin set before his model judge, Hercules.[94] But that may not be something that the computer program we have named Hercules can do. If we are right about AI being unable to make substantive moral judgments, Hercules, as AI, cannot perform all of the tasks it needs to perform to select between competing legal interpretations, at least according to Dworkin. Ultimately, substantive *moral* judgments are necessary for Hercules to go about its business. And moral judgments may be uniquely *human* judgments.

We might imagine that, in the not so distant future, computers will be able to assess "fit" more quickly and accurately than human beings can, just as they can beat us at chess or Go. But we have discussed reasons to doubt that computers will be able to assess "justification" more accurately than human beings or, indeed, to doubt that they will be able to assess "justification" at all.[95] The above reasoning leads us to two propositions: (1) substantive moral judgments may sometimes be necessary to say what the law is; and (2) human beings alone may be capable of making those substantive moral judgments. These propositions taken together can shape the role we assign AI in legal interpretation. If we can determine when legal interpretation requires substantive moral judgment, we will know when human beings have a special role to play in our legal system. There is also an intriguing possible twist. If we are right about the extraordinary potential of AI, then a third point may also be true: (3) computers may be capable of assigning odds to the various moral judgments human beings might make, even if they are not capable of making substantive moral judgments. These three points take us to the intersection of AI and jurisprudence, the subject to which we now turn.

III. JURISPRUDENCE: MORALITY AND LEGAL INTERPRETATION

The discussion above concerning computers and legal interpretation can motivate an inquiry into jurisprudence in two ways. First, it gives jurisprudence additional relevance and urgency. We have

93. *Id.*

94. *Id.*

95. But note computers may be able to predict how human beings would assess justification. *See* Part III.D (discussing some potential implications of that possibility).

a new reason to resolve the longstanding dispute about the role morality plays in legal interpretation. Doing so can help us understand and define the functions AI can—and cannot—perform in legal interpretation.

Reflecting on AI can motivate jurisprudential inquiry in a second way. We may have a strong intuition that we should circumscribe the role of computers in legal interpretation—that we should reserve a place for human beings. Attention to the role of morality in legal interpretation can help us understand and justify that intuition. We may sense that in some circumstances we want legal interpreters to exercise substantive moral judgment—and we may believe that is something computers cannot do. Our view of the proper part for computers to play in legal interpretation, then, may inform our views of the role of morality in legal interpretation.

These points provide context for exploring a novel jurisprudential position. I have begun to develop an argument in various publications—some of them co-authored by Manuel Vargas, a philosopher—called "Legal Dualism."[96] It holds that the best account of the nature of the law varies with the purpose of legal interpretation.[97] When a legal interpreter seeks merely to describe the law or to predict how others will interpret it, Legal Dualism suggests that there is no need to make moral judgments in saying what the law is.[98] In other words, legal positivism provide the best account of the nature of law in these circumstances.[99] However, at other times, legal interpreters seek moral guidance from the law. When they do, Legal Dualism holds, legal interpreters often—perhaps always—need to make substantive moral judgments.[100] So natural law provides the best account of the nature of law under these conditions.[101]

Legal Dualism offers a solution to the primary dispute in jurisprudence. It also provides a way of mapping AI's potential and limitations in legal interpretation. If, as suggested above, computers

96. *See* Vargas & Davis, *American Legal Realism and Practical Guidance*, *supra* note 19; Davis & Vargas, *Legal Dualism, Naturalism, and the Alleged Impossibility of a Theory of Adjudication*, *supra* note 19.

97. *Id.*

98. *Id.*

99. *Id.*

100. *Id.*

101. *Id.*

will soon outstrip people in describing the law and predicting how it will be interpreted, they may displace the human interpreters charged with undertaking those tasks. On the other hand, if computers cannot make substantive moral judgments—if they must be told what ultimate ends to pursue and cannot choose those ends on their own—then Legal Dualism suggests when people should retain responsibility for legal interpretation.

Of course, this reasoning matters only if legal interpreters have an obligation at times to look to the law for moral guidance. There are reasons to think they do. Some legal scholars believe that the law generally has moral force. Even for those who doubt that general claim, there may be circumstances that they think impose an obligation on legal interpreters to follow the law.[102] One likely candidate would be judges when they rule on matters of law. They must select from a range of possible legal interpretations in deciding the case before them.[103] Another candidate would be lawyers in some circumstances—including, arguably, when offering advice about the law to those charged with developing the software for self-driving cars to make life and death decisions.[104]

Another jurisprudential issue is also worth noting. If we think AI will ultimately be able to outperform people in all tasks of description or prediction, that superiority could extend to certain questions about morality as well. In particular, AI might be better able to describe the sets of moral beliefs that people hold—or profess, or act on (which are not necessarily the same)—at least under particular circumstances.[105] But that does not mean that computers will be able to exercise substantive moral judgment. That distinction suggests an area for potential jurisprudential inquiry, one that actually arose long ago when Lon Fuller and H.L.A. Hart engaged in a famous colloquy about the role of morality in legal interpretation.[106] In other words, we might

102. *See* Davis, *Legal Dualism*, *supra* note 19, at 21-26 (discussing possibility of law imposing moral obligations).

103. *Id.* at 22.

104. *Id.* at 23-25.

105. *But see* Part II.C (discussing the potential difficulties for AI in making these sorts of predictions in legal interpretation).

106. *See* H.L.A. Hart, *Positivism and the Separation of Law and Morals*, 71 HARV. L. REV. 593 (1958); *see also* Lon L. Fuller, *Positivism and Fidelity to Law—A Reply to Professor Hart*, 71 HARV. L. REV. 630 (1958).

distinguish between descriptive and predictive claims *about morality*, on the one hand, and *substantive* moral judgments, on the other. For some purposes—perhaps in the context of defining the appropriate scope of AI—we might distinguish theories about the nature of law that permit descriptive or predictive judgments about morality in legal interpretation from those that permit substantive moral judgments. That approach could supplement other ways in which scholars distinguish legal positivism from natural law.

Part III explores these points. Part III.A frames the central historical debate in jurisprudence, that is, between natural law and legal positivism. That central debate is about the role of morality in legal interpretation—a debate with implications for the potential outer boundary of the functions AI can perform.

Part III.B explains how Legal Dualism offers a possible solution to that debate, one in which natural law and legal positivism each provide an appropriate account of the nature of law within a properly defined setting. Legal Dualism may also explain why human beings play an essential role in legal interpretation when the law serves as a source of moral guidance. It may thereby provide a way of circumscribing the role of AI in legal interpretation, including as it pertains to guiding driverless cars.

Part III.C then suggests some circumstances in which the law should serve as a source of moral guidance. One likely example is when judges (or juries) resolve legal disputes. Others likely arise when a self-driving car decides which lives to spare and which to sacrifice, and when a corporation decides how an autonomous car should be programed to go about making life and death decisions.

Part III.D suggests ways in which reflecting on technological advances may improve our understanding of jurisprudence. It notes that our concerns about AI—and their potential role in legal interpretation—may help to justify Legal Dualism. If Legal Dualism can explain why we might want to limit the role of AI in our legal system—and does so in a way that makes sense of our intuitions—then that suggests a reason to endorse Legal Dualism. Part III.D also explores a parallel between the potential for computer interpretation of the law and computer assessment of morality: computers may be able to describe morality and predict the moral judgments people will make, but they may be unable to make substantive moral judgments. That suggests an interesting alternative way to define the different school of jurisprudential thought,

a way that derives from contemplating the role of AI in legal interpretation. For some purposes, the best way to distinguish legal positivism from natural law may not be based on the role of morality in legal interpretation but rather on the role of *substantive* moral judgment in legal interpretation. In this way, we can see not only how jurisprudence can inform cutting edge issues in legal interpretation but vice-versa as well. The prospect of—or, if you prefer, a thought experiment about—AI interpreting the law may cast new light on longstanding jurisprudential debates.

A. Jurisprudence: Defining the Nature of Law by the Role of Morality

If the discussion above about Hercules is correct, AI's role in legal interpretation may be circumscribed by the need to make substantive moral judgments in saying what the law is. This point dovetails nicely with the central dispute in jurisprudence for the past half century: morality's role in legal interpretation.[107] The major schools of thought in jurisprudence are defined by the role they believe morality plays in law. The primary rift lies between legal positivism and natural law.[108]

Legal positivism draws a distinction between law and morality.[109] There are various branches of legal positivism. Exclusive (or hard) legal positivists claim that morality plays no role in saying what the law is.[110] To be sure, exclusive legal positivists acknowledge that morality can and should figure in creating the law—when a legislature enacts a law, for example, or perhaps when a judge makes new law. But they believe that is a separate issue from determining what the law is at a given time.[111]

Inclusive (or soft) legal positivism, in contrast, accepts a possible—and contingent—role for morality in the law.[112] One typical way to define inclusive legal positivism is by the social facts thesis. It holds that the content of the law depends ultimately only on social facts.[113]

---

107. *See* Hershovitz, *supra* note 8, at 1162; *see also* Davis, *Legality*, *supra* note 5, at 61-63.

108. *Id.*

109. *See* SHAPIRO, *supra* note 5, at 273-74.

110. *Id.* at 271.

111. *Id.*

112. *Id.* at 269-70.

113. *Id.* at 269.

However, those ultimate social facts may make morality relevant to particular legal judgments, incorporating morality into the law.[114]

A third approach has been called normative (or ethical) legal positivism.[115] That version of positivism is in a sense the inverse of inclusive legal positivism. Normative positivism allows morality to play a part in ultimate (or foundational) judgments about the law, but only if those moral judgments lead to the conclusion that legal interpreters should not make additional moral judgments in saying what the law is.[116] Justice Antonin Scalia arguably belonged to this school of thought.[117]

Non-positivists—sometimes called natural lawyers—tend to resist the distinction between law and morality. They view moral judgments as at times necessary in making legal judgments. Ronald Dworkin espoused an "interpretivist" view that integrated substantive moral judgments with descriptive judgments in legal interpretation.[118] Lon Fuller had a similar view in some regards, although he tended to focus on a special class of moral values that arise distinctively in legal systems—on what he called the "inner morality" of law.[119]

The position one takes on these jurisprudential issues can have implications for the scope of AI in legal interpretation, particularly if one accepts that AI can learn to make all of the descriptive and predictive judgments that human beings make about law and do a better job at them, but believes that AI cannot make substantive moral judgments. If so, then human beings may have a special role to play in legal interpretation to the extent—perhaps only to the extent—that natural law provides the best account of the nature of law in some

114. *Id.* at 270-71.

115. Jeremy Waldron, *Normative (or Ethical) Positivism*, *in* HART'S POSTSCRIPT: ESSAYS ON THE POSTSCRIPT TO THE CONCEPT OF LAW 411 (2001).

116. Davis, *Legal Dualism*, *supra* note 19, at 8.

117. *Id.* at 12-14. Non-positivists (or natural lawyers) have similar differences of view, although they are usually not as clearly defined. Lon Fuller, for example, gave morality a relatively cool embrace. He tended to focus on the "inner morality of the law," that is, on moral values that were distinctive to the legal realm. *See* Fuller, *supra* note 106, at 650, 659-60; *see also* LON L. FULLER, THE MORALITY OF LAW (1964). Ronald Dworkin, in contrast, drew on morality much more broadly. *See, e.g.*, RONALD DWORKIN, JUSTICE FOR HEDGEHOGS (2011).

118. *See id.*

119. *See* Fuller, *supra* note 106, at 650, 659-60; *see also* FULLER, *supra* note 117.

circumstances, that is, to the extent that legal interpretation requires substantive moral judgments.

The most obvious example is exclusive legal positivism. Exclusive positivists believe that no substantive moral judgments are necessary to say what the law is.[120] If computers become better at making purely positive judgments about the law than human beings and if exclusive legal positivists are right about the nature of law, then AI would seem capable of displacing human beings as legal interpreters. To be sure, even many legal positivists contemplate some role for morality in adjudication. Scott Shapiro, for example, has claimed that the legal interpretation should require an assessment only of social facts, but he acknowledges that making new law—which judges sometimes do—and applying the law—which judges often do—can require moral judgments.[121]

Similarly, inclusive legal positivists accept that the law sometimes incorporates moral judgments, even though the content of the law—including whether it requires or permits moral judgments—is *ultimately* a matter only of social fact.[122]

Even if legal positivists are right about the nature of law, then, human beings might have a role to play in legal interpretation. But that role likely would be limited. There will come a day when AI could—and presumably should—make all of the purely descriptive and predictive judgments necessary to interpret the law.[123] It would also

120. SHAPIRO, *supra* note 5, at 269.

121. *See* Davis, *Legality*, *supra* note 5, at 76-77 (discussing SHAPIRO, *supra* note 6, at 274-76).

122. In other words, exclusive and inclusive legal positivists agree on the Social Fact Thesis. SHAPIRO, *supra* note 5, at 273.

123. There may be parts of the law that are neither susceptible to analysis by AI nor involve substantive moral (or other value) judgments. Brian Leiter has offered a particularly well-developed view along these lines. *See* Brian Leiter, *Heidegger and the Theory of Adjudication*, 106 YALE L.J. 253, 253 n.3 (1996). For a response to Leiter's view on this point see Davis & Vargas, *Legal Dualism, Naturalism, and the Alleged Impossibility of a Theory of Adjudication*, *supra* note 19.Stanley Fish provides a valuable discussion of legal "craft." *See, e.g.*, STANLEY FISH, DOING WHAT COMES NATURALLY: CHANGE, RHETORIC, AND THE PRACTICE OF THEORY IN LITERARY AND LEGAL STUDIES (1989); Stanley Fish, *Dennis Martinez and the Uses of Theory*, 96 YALE L.J. 1773 (1987); Dan M. Kahan, *The Supreme Court 2010 Term: Foreword: Neutral Principles, Motivated Cognition, and Some Problems for Constitutional Law*, 125 HARV. L. REV. 1, 27-28 (2011) (discussing "craft norms"). Also, like Leiter, Jack Balkin relies explicitly on the work of Heidegger. *See, e.g.*,

identify and frame the moral judgments necessary to resolve a legal dispute. Human beings would then be called upon to fill in the missing pieces. So we might imagine that human judges could come to play the sort of narrow role that juries play in our legal system today—resolving defined and circumscribed issues within a legal system largely run by AI. And lawyers might play an analogous part in other settings.

These matters of legal theory are not the sole province of academics. Consider Chief Justice John Roberts' opening statement at his confirmation hearing. He famously compared judges to home plate umpires: "And I will remember that it's my job to call balls and strikes and not to pitch or bat."[124] Metaphors can be interpreted in many ways. So it is dangerous to read too much into this claim. But one way to understand it is as declaring the Chief Justice a legal positivist, likely an exclusive legal positivist or a normative legal positivist. Home plate umpires, after all, arguably make a purely descriptive assessment: was a pitch a ball or a strike? They, ideally, do not make their own value judgments in defining the strike zone but rather mechanically follow the rule embodied in the Major League Baseball rule book.[125]

Confirmation that the Chief Justice subscribes to legal positivism lies in his statement that he would decide cases based on "the rule of law" and his declaration that he had "no agenda."[126] Again, these words are somewhat vague. But one plausible way to understand them is as eschewing moral judgments. If Justice Roberts did not anticipate

---

Jack M. Balkin, *Understanding Legal Understanding: The Legal Subject and the Problem of Coherence*, 103 YALE L.J. 105, 159 n.110 (1993).

124. *Roberts: 'My Job is to Call Balls and Strikes and not to Pitch or Bat,'* CNN (Sept. 12, 2005, 4:58 PM), http://www.cnn.com/2005/POLITICS/09/12/roberts.statement/.

125. Major League Baseball Rule 2.00 defines the strike zone as "that area over home plate the upper limit of which is a horizontal line at the midpoint between the top of the shoulders and the top of the uniform pants, and the lower level is a lien at the hollow beneath the kneecap" and is determined by "the batter's stance as the batter is prepared to swing at a pitched ball." The tenor of the Chief Justice's remarks suggests that no substantive value judgment by an umpire is necessary to determine the content of this rule. One could, of course, argue the contrary. But that would seem to involve a critique of the Chief Justice's jurisprudential perspective rather than an effort to determine how the Chief Justice himself views the nature of law.

126. *Roberts: 'My Job is to Call Balls and Strikes and not to Pitch or Bat,' supra* note 124.

making moral judgments at all in interpreting the law, he qualified as an aspiring exclusive legal positivist.[127]

Understanding the law as exclusive legal positivists do—as not requiring moral judgment—suggests legal interpretation is a task that computers may soon be able to do better than human beings. Here we can extend the Chief Justice's metaphor. We might wonder why we still rely on human umpires for calling balls and strikes. Why not assign *that* job to a computer? Baseball fans have seen on television that computers appear able to distinguish balls and strikes more accurately and more reliably than human beings in real time. That is presumably why sports networks use computers to analyze umpires' calls after the fact. And if computers are not more accurate and reliable yet, they likely will be in the near future. Indeed, some sports already rely on computers to make key judgments initially or to review human judgments.[128] Think of line calls in tennis.[129] It seems the main reason not to rely on computers is some sort of bathetic nostalgia for tradition. Relying on computers as umpires or referees, as the British might say, "just isn't cricket."[130]

In the context of entertainment—and, no matter how seriously we take them, sports are after all a form of entertainment—continuing to rely on human umpires or referees is understandable, even if they are more susceptible to bias and less reliable and less accurate than machines. After all, machines might also be better than human *players*, but having a robot pitch or bat would ruin the game. If we prefer human

127. If instead he planned to make only ultimate or foundational moral judgments—about, for example, the need for unelected judges in a democracy to refrain from making substantive moral judgments in legal interpretation, perhaps for reasons sounding in democratic theory, as Justice Scalia arguably did, Davis, *Legal Dualism*, *supra* note 19, at 12-13—then he is a normative positivist. That distinction does not matter for purposes of the discussion in the text.

128. Tom Perrotta, *Hawk-Eye is Here to Kill Tennis*, WKLY. STANDARD (Mar. 19, 2018, 12:05 PM), https://www.weeklystandard.com/tom-perrotta/hawk-eye-is-here-to-kill-tennis.

129. *Id*.

130. Except it is, ironically. Cricket is one of the sports that has adopted technology to ensure accuracy. *See* James McKern, *Talking Points from Day Two of the Second Ashes Test in Adelaide*, NEWS (Dec. 4, 2017), https://www.news.com.au/sport/cricket/the-ashes/talking-points-from-day-two-of-the-second-ashes-test-in-adelaide/news-story/55a4a5164fc406b334cb59af186cc4c4. That controversy has resulted is unsurprising, even inevitable. *Id.*

umpires and referees as well we are free to indulge that inclination, even if we regularly experience unnecessary mistakes as a result.

The same may not be true for the law. If computers surpass human beings at the purely descriptive aspects of legal interpretation—if they ultimately prove superior at calling "balls and strikes"—and if that is all that legal interpretation entails, perhaps we should allow computers to take over. They can become our attorneys—and even our judges. Nostalgia has some value, but so too do speed, accuracy, and, ultimately, "the rule of law," to quote the Chief Justice. It is thus not surprising, as discussed above, that Max Tegmark's suggestion that "the legal process can be abstractly viewed as computation"—by which he seems to mean a mechanical, if complex, process—led him to conclude that robojudges would appear superior in many regards to human judges.[131] In theory, robojudges might be less biased, not susceptible to fatigue, and capable of amassing vast knowledge.[132] If legal interpretation is just computation, we may have a moral obligation to get past our nostalgia and turn our legal system over to AI. Let Hercules do what it does better than human judges. Ludditism would seem a poor competitor to the rule of law.

But what if legal interpretation does not entail mere computation but requires something more? What if it sometimes requires substantive moral judgments? And what if we cannot define the goals of morality in a way that enables AI to make substantive moral judgments? In those circumstances, we might have reason to retain a role for human beings in saying what the law is.

B. Legal Dualism: Resolving the Central Jurisprudential Debate and Circumscribing the Role of Computers

Our inquiry into AI's potential role in legal interpretation, then, lends new importance and timeliness to the longstanding debate between legal positivism and natural law. One possible solution to the debate is Legal Dualism. It suggests that the best account of the nature of law depends on the goals of a legal interpreter.

Legal Dualism focuses, in particular, on the different ways legal interpreters can best contend with legal indeterminacy. At least since

131. TEGMARK, *supra* note 1, at 105.

132. *Id.*

the work of the Legal Realists in the late nineteenth and early twentieth centuries, most legal theorists accept that the law has a significant degree of indeterminacy. There are often multiple legal interpretations consistent with authoritative sources of law on a legal issue.[133] Legal Dualism proceeds from the valuable insight that how a legal interpreter should deal with that indeterminacy depends on her goals in interpreting the law.

Some legal interpreters can—and should—leave legal indeterminacy intact. A person simply describing the law can identify the different possible legal interpretations and explain how they can be reconciled with authoritative sources of law.[134] If she were to eliminate that indeterminacy—if she were to describe only what she thinks is the best account, recognizing that there is sincere and reasonable disagreement—her description is likely to be incomplete and suffer as a result.

The same is true for someone predicting how the law might be interpreted. To the extent it is appropriate to analyze the law for purely prudential reasons—for example, assessing potential legal liability in acquiring a particular business—the best strategy may be to assign odds to different potential legal interpretations and assess their financial consequences. Applying only one of several possible legal interpretations to an important issue could lead to poor decision-making. The goal is not to determine the best legal interpretation—in the view of any given interpreter—but rather to make a sound financial investment recognizing uncertainty, acknowledging that different legal outcomes are possible, and weighing the odds and consequences of each one.

Legal interpreters looking to the law for moral guidance are differently situated. They often must resolve the indeterminacy in the law if they are to take the law into account in deciding what they should do.[135] Consider a judge. Let's assume that the judge has some moral obligation to attempt to follow the law in rendering a decision. The judge may recognize that multiple interpretations of the relevant law are available.[136] But in ruling she may have to choose only one. The

133. *See* Davis, *Legal Dualism*, *supra* note 19, at 33.

134. *Id.* at 18; Davis, *Legality*, *supra* note 5, at 92.

135. *Id.* at 92-23.

136. *Id.*

judge cannot simply make a prediction about how she is likely to rule.[137] Nor would she seem to fulfill her duties if she were to rule based on a simple prediction about how other judges—perhaps appellate judges—would likely interpret the law. Instead, she may have to resolve any indeterminacy in the law using her best judgment and, according to at least one plausible theory, she can do so by taking into account the relative fit of each possible interpretation to authoritative sources of law and its relative justification.[138] In other words, she may have to make substantive moral judgments—likely those framed by the relevant legal rules and standards—to render the law sufficiently determinate and reach a conclusion.

This is not the setting to develop the full argument for Legal Dualism, to explore the criticisms that may be leveled against it, or to assess whether Legal Dualism has adequate responses to them. Those issues will tend to be general and lack any special significance for AI's role in ethics. But, as noted above, there is one point about Legal Dualism that has particular relevance to the current discussion. It can help to explain and justify limiting AI's role in legal interpretation and, perhaps, rejecting the notion of robojudges.

Let us assume four propositions are true. First, the law is sufficiently indeterminate that multiple legal interpretations are available on many legal issues. Second, judges—and some other legal interpreters—should look to the law for moral guidance. Third, when legal interpreters seek moral guidance from the law, and the law is indeterminate absent moral judgment, interpreters should exercise substantive moral judgment to render it sufficiently determinate to help guide them. Fourth, AI is not capable of making substantive moral judgments; it needs human beings to choose the ultimate ends that AI will pursue. If we accept those propositions, Legal Dualism provides a way to understand and justify limiting AI's role in legal interpretation. We should rely on human legal interpreters—not AI—when the law is indeterminate absent moral judgment and serves as a source of moral guidance. When is that likely to occur?

137. HART, *supra* note 77, at 124-54.

138. *See supra* notes 70, 90 and accompanying text.

C. Application of Legal Dualism: Law as a Potential Source of Moral Guidance

According to Legal Dualism, moral judgment is required for legal interpretation only when the law serves as a source of moral guidance. At issue, then, is whether the law ever plays that role. Some might think it does not. Oliver Wendell Holmes, Jr., for example, famously defined the law as "prophecies of what courts will do in fact, and nothing more pretentious."[139] He also developed his "Bad Man" theory of the law, which suggests that the only concerns to which law gives rise are prudential, not moral. As discussed above, computers may soon be far more effective at predicting legal rulings—at offering "prophecies"—than human beings. If so, and if one subscribes to Holmes' brand of legal positivism as described above, there may be no special role for human beings in interpreting the law, even if one accepts Legal Dualism.

But there is reason to question Holmes' positivism. As H.L.A. Hart pointed out, predicting judicial rulings does not provide particularly useful guidance to judges.[140] Is a judge supposed to rule by predicting how she will rule? That seems hopelessly circular. It also seems particularly plausible that judges have some moral obligation to abide by the law in their rulings.[141] They take an oath (or affirmation) to do so.[142] Moreover, they have special powers and arguably special moral dispensations based on that possible moral obligation. We likely would not empower judges to make momentous decisions if we did not think them morally bound by the law. And we might assess harshly the way they exercise power—throwing people in prison for decades, stripping them of millions of dollars—if we thought they were individually morally responsible for their actions rather than permitted by their role to act in ways that would otherwise be morally suspect.[143] And we

139. Oliver W. Holmes, Jr., *The Path of the Law*, 10 HARV. L. REV. 457 (1897).

140. *See* HART, *supra* note 77, at 124-54.

141. This may be true even if other moral considerations may at times have great weight.

142. 28 U.S.C.A. § 453 (Westlaw through Pub. L. No. 115-231).

143. *See* Davis, *Legal Dualism*, *supra* note 19, at 20-22 (discussing this notion of reciprocity—of the law protecting people from moral accountability for their actions when they operate within their institutional roles in the legal system). Relevant to this issue is Hart's distinction between the reason for obeying the law and

think litigants have special grounds to complain when a judge's ruling is not only unfavorable to them but is also "lawless." So judges likely have a moral obligation at least to take the law into account in deciding what they should do. To that extent, Legal Dualism can help explain why robojudges should not displace human judges.

The analysis is more complicated when it comes to attorneys. One of the central disputes in legal ethics is whether attorneys can operate essentially as legal positivists—offering whatever options and making whatever legal arguments best meet their clients' needs or desires—or whether attorneys are obligated to exercise some independent judgment in interpreting the law. That debate often focuses on whether our legal system is best understood as embodying a strong form of the adversarial system.[144] This is not the place to take a position on that debate.[145] Note, however, that many well respected scholars are skeptical about justifying aggressive attorney behavior in the name of the adversarial system—at least in some settings. If one shares that skepticism, and also accepts Legal Dualism, then one may conclude that attorneys should give their independent views on how best to interpret the law and legal obligations, and not just predict potential outcomes as would Holmes' Bad Man. According to Legal Dualism, that means attorneys sometimes have an obligation to exercise moral judgment in interpreting the law. When they do, the limitations of robojudges may extend to roboattorneys. Neither form of AI may be able to make the substantive moral judgments required to act in an ethical manner. Hercules—brilliant as it is—has a limited ability to play the role of lawyer, just as it has a limited ability to preside as a judge.

If attorneys have an obligation to exercise independent legal judgment in some circumstances—and to exercise *moral* judgment in doing so—a likely example of when that obligation arises would be in

---

for abiding obeying the commands of a man carrying a gun. *See* HART, *supra* note 77, at 6, 20, 90.

144. *See*, *e.g.*, W. BRADLEY WENDEL, LAWYERS AND FIDELITY TO LAW (2010); TIM DARE, THE COUNSEL OF ROGUES?: A DEFENCE OF THE STANDARD CONCEPTION OF THE LAWYER'S ROLE (2009); DANIEL MARKOVITS, A MODERN LEGAL ETHICS: ADVERSARY ETHICS IN A DEMOCRATIC AGE (2008); DAVID LUBAN, LEGAL ETHICS AND HUMAN DIGNITY (2007); DEBORAH RHODE, IN THE INTERESTS OF JUSTICE (2000); WILLIAM H. SIMON, THE PRACTICE OF JUSTICE: A THEORY OF LAWYERS' ETHICS (2000).

145. *See*, *e.g.*, LUBAN, *supra* note 144; RHODE, *supra* note 144; SIMON, *supra* note 144.

assessing the legal standards that inform how to program self-driving cars. Consider a related historical precedent: the Ford Pinto case. Ford designed its Pinto—including the location of the gas tank and related safety measures—in a way that created an unusually high risk of personal injury in a rear-end collision. *Grimshaw v. Ford Motor Company* addressed a resulting lawsuit, wherein an injured motorist sought not only compensation for injuries but also punitive damages.[146] Ford argued that California law required malice to impose punitive damages and that because Ford did not want to harm anyone—at most it risked harm as a business decision—it could not have possessed the requisite malice. The California Court of Appeal disagreed in a way that casts some light on a corporate actor's duties to look beyond legal consequences as would Holmes' Bad Man.

The California appellate court rejected the notion that Ford had to "intend to harm a particular person or persons" to act with malice.[147] Instead, it held that the word "malice" encompasses "conduct evincing callous and conscious disregard of public safety by those who manufacture and market mass produced articles."[148] Further, it defined "callous and conscious disregard of public safety" to include a decision by a car manufacturer "to treat compensatory damages as a part of the cost of doing business rather than to remedy [a] defect."[149] One way to interpret the court's reasoning is as imposing on Ford an obligation to abide by the law designed to protect the public safety and not to treat the risk of legal liability as merely a financial consideration in maximizing its profits, a la Holmes' bad man. Deliberately violating that obligation—because on net paying compensatory damages would be profitable—was sufficient to support punitive damages.[150]

146. Grimshaw v. Ford Motor Co., 119 Cal. App. 3d 757, 771-72 (1981).

147. *Id.* at 809.

148. *Id.* at 810.

149. *Id.*

150. Note that violation of a legal right is required for a punitive-damage claim; there is no freestanding cause of action for punitive damages. So, the fact that Ford's conduct violated the law—and was not just that it was arguably morally reprehensible—was necessary to the court's reasoning. If a corporate actor put members of the public at risk in a way that did not violate a legal standard—for example, if its conduct was deemed not to constitute a tort—there would be no claim for punitive damages. Punitive damages require an underlying legal violation. So, the

Of course, the California appellate court's opinion does not resolve whether Ford—and the lawyers who advised it—had a moral obligation to take the law into account in deciding how to design their cars. A judicial opinion indicating that there is a duty to follow the law—and not to treat legal sanctions as mere incentives—does not resolve the matter of whether the law in fact creates moral obligations. The question simply repeats itself. Does Ford have a moral obligation to follow the judicial mandate expressed in imposing punitive damages? Or is it morally permissible for Ford to treat punitive damages themselves—in addition to compensatory damages—as merely the "cost of doing business," as would Holmes' Bad Man?

But *Grimshaw* is at least suggestive. It indicates that the law itself considers the relevance of a legal rule to be more than prudential and to give rise to obligations. And it implies that Ford did something morally wrong—that it acted with "malice," a term freighted with moral and not just legal significance—not only by putting people at risk but also by violating their legal rights.

Further, legal interpretations that govern a car manufacturer's decisions are not subjected to the rigors of the adversarial process before they have grave consequences. When lawyers are airing their arguments in open court—and a judge has an opportunity to scrutinize them and select among them—lawyers are arguably relatively free to take aggressive legal positions.[151] The judicial system serves, if you will, as a safety net. The judge arguably is charged with exercising independent legal—and perhaps moral—judgment. That can liberate the attorneys.[152] But the same is not true for a car manufacturer whose decisions may cost many lives—and may preserve others—before the judicial system has any realistic prospect of evaluating the car manufacturer's analysis and possibly ordering it to modify its conduct. Under these circumstances—where an actor in the world and its attorneys make decisions that have consequences long before they are assessed by an impartial tribunal—the argument is particularly strong that attorneys have an obligation to render independent judgment in

violation of a legal standard—not just a moral standard—seems to be doing some work in this example.

151. Davis, *Legal Dualism*, *supra* note 19, at 22-25.

152. *Id.* at 24-25.

saying what the law is.[153] Those attorneys cannot rely on the corrective effects of the adversarial system to the same extent as lawyers arguing in court.

The kind of obligation addressed in the Ford Pinto case to take the relevant legal standard into account in deciding how to act morally would extend naturally to programming self-driving cars. Like the placement of a gas tank in the body of a car, programming decisions will have life-and-death implication. If a car company makes the decision in a way that treats human injury and death in violation of people's legal rights as a mere cost of doing business, punitive as well as compensatory damages could result.

Moreover, as discussed above, the way in which a car makes those decisions—at least if it relies on AI—may be very difficult to discern, rendering causation difficult to trace. The calculation that a self-driving car made in choosing a course of conduct may be obscure and abstruse—and thus difficult to assess even in hindsight.[154] As a practical reality, the computer programmers in the backroom of an autonomous-car manufacturer—and the lawyers who advise them—will be rendering crucial decisions subject to limited scrutiny and with profound implications. These are just the sorts of circumstances in which we might expect lawyers and their clients to have a particularly heavy obligation to take into account their legal and other moral obligations—assuming legal obligations create moral obligations. Any adversarial process is likely to have at most a belated effect on crucial decisions. Even then, its prospects for discovering what happened and who is responsible may be poor.[155] Like the lawyer advising a client on how to destroy evidence to get away with murder, we are likely to be skeptical that an attorney can appropriately advise autonomous-car manufacturers to take actions violating relevant legal standards just because the manufacturer is likely to "get away with it." Attorneys advising clients that program self-driving cars are particularly likely to have an obligation to exercise independent judgment in determining what the law requires and not just to lay out the potential financial consequences of different choices. If Legal Dualism provides a

153. *Id.* at 24.

154. *See supra* Part III.C.

155. *See* Davis, *Legal Dualism*, *supra* note 19, at 22-25; *see also* WENDEL, *supra* note 143, at 187-92.

persuasive account of the nature of law and legal interpretation, that independent legal judgment would at times entail *moral* judgment.

D. Jurisprudence Informing AI; AI Informing Jurisprudence

1. Moral Judgments as Circumscribing AI's Role in Legal Interpretation

We have seen that lessons from jurisprudence about the nature of law and legal interpretation may provide guidance about AI's potential and its limitations. Robointerpreters may soon outstrip human beings in providing descriptive and predictive—perhaps even persuasive—accounts of the law. But it is far less clear that even Hercules can be programmed to make the substantive moral judgments that may sometimes be necessary to say what the law is. If Hercules cannot be so programmed, and if legal interpretation sometimes requires moral judgments, there may be a line that AI cannot cross in the foreseeable future, even if its technical capacities continue to increase at an extraordinary rate. There may a difference in kind—and not just degree—between the judgments Hercules can make and some of the judgments—the moral judgments—that are necessary at times for legal interpretation.

As a result, the role that Hercules can play may depend in part on the best understanding of the nature of law. If exclusive legal positivism captures the nature of law, AI may soon be able to perform legal interpretation more effectively than human beings. After all, saying what the law is requires only an assessment of social facts. If we want accurate, reliable, and efficient interpretation of the law, it may then make sense for us to delegate the great bulk of the work in legal interpretation to computers. They may well be better at it than us, capable of completing most of the tasks of judges and lawyers. Of course, there may still be some role for human beings. We may be uniquely capable of making the moral or other value judgments necessary to make new law. And perhaps the same is true in *applying* the law, as opposed to saying what it is.[156] But AI should be able to do

156. *See, e.g.*, SHAPIRO, *supra* note 5, at 141-47 (acknowledging that moral judgments may be necessary at times in adjudication, even if not in legal interpretation).

the rest and then to identify and frame the assessments for human judges or lawyers to make.[157]

Inclusive legal positivism, in contrast, could preserve a larger role for human beings. It accepts the possibility that the law may incorporate moral judgments. To the extent the law does, human beings may have some special role to play in saying what the law is. That said, much like with exclusive legal positivism, AI could still do a great deal of legal interpretation for us, framing the issues that it has to leave open so that they can be resolved by human beings capable of making substantive moral judgments.

Legal Dualism, in contrast, suggests that moral judgments are necessary for legal interpretation in a range of settings. In particular, according to Dualism, a legal interpreter must make moral judgments when she has a moral obligation to take the law into account in deciding how to act. Legal Dualism can thus help us recognize when Hercules should not engage in legal interpretation. We might generally conclude that we should not hire robojudges instead of human judges—notwithstanding the perceived advantages of robojudges in some regards[158]—and, at least in some circumstances, that we should not displace human attorneys with roboattorneys.[159] More generally yet,

157. Roughly the same analysis would seem to apply to normative (or ethical) positivists. Recall that according to normative (or ethical positivists), only a foundational (or ultimate) moral judgment is required to adopt positivism. The programmers of AI can make that judgment. For these purposes, normative (or ethical) positivism would seem to collapse into exclusive legal positivism. Of course, it is possible that in some circumstances the foundational (or ultimate) judgment would lead us to abandon positivism, that normative (or ethical) positivism keeps that possibility alive, and that normative (or ethical) positivism could therefore diverge from exclusive legal positivism in some circumstances. Justice Scalia, for instance, once famously said he would be a "faint-hearted" originalist if flogging was originally not considered cruel and unusual punishment. Antonin Scalia, *Originalism: The Lesser Evil*, 57 U. CINN. L. REV. 849, 864 (1989). He later retracted that position. Ilya Somin, *Justice Scalia Repudiates "Faintheated" Originalism*, VOLOKH CONSPIRACY, http://volokh.com/2013/10/07/justice-scalia-repudiates-fainthearted-originalism/.

158. TEGMARK, *supra* note 1, at 105-06.

159. Whether human attorneys in fact act with the independent judgment they should—rather than simply tell their clients what they want to hear—is a somewhat separate, although related, question. So is the question of whether in the future we may want to require—or incentivize—clients to rely on independent, human legal counsel rather than on robolawyers in making decisions with significant legal

we can recognize that jurisprudence may be able to inform cutting-edge ethical issues in regard to computer learning.

The converse is true as well. Contemplation of developments in technology and the ethical issues to which they give rise may inform jurisprudence. If, for example, we doubt that robojudges—or even robolawyers—are a good idea, and if one of the reasons why we harbor those doubts is that we suspect judges and lawyers should exercise moral judgment, that gives us a reason to find Legal Dualism attractive, in addition to any other arguments in its favor. If Legal Dualism can explain and justify our resistance to giving Hercules too big a role in our legal system, that is a point in Dualism's favor.

2. Different Kinds of Moral Judgments

There may be other ways in which attention to technological advances can inform jurisprudence. Consider an issue noted briefly above. AI may not only be able to make certain judgments about the law but also certain judgments about morality. Morality, like law, is in part a social practice. There are interesting social facts about morality. Even if AI cannot make substantive moral judgments, it may be able to describe, as a matter of social fact, the moral beliefs that people hold, to predict the moral judgments people will make, and even to identify moral arguments that people will find persuasive.[160]

Note the parallel role that Hercules could play in performing legal analysis. Legal interpreters—including artificial ones—can make different kinds of claims: descriptive, predictive, what one might call persuasive, and what one might call substantive. In other words, a legal interpreter can seek to describe the law, to predict how others will interpret the law, or to frame an analysis of the law in a persuasive way (which involves a kind of prediction about how that person will respond). Or the legal interpreter can make what might be called a substantive legal judgment, attempting to provide the best legal interpretation she can, including, at least according to Legal Dualism, in some cases making the moral judgments necessary to render the law

implications. Can a computer, one might ask, engage in the unauthorized practice of law? These issues are beyond the scope of this Article.

160. Note, however, that AI might have difficulty disentangling moral (or other value) judgments from the other motivations for the decisions people make. *See supra* Part II.C.

determinate and, thus, provide moral guidance. The same kinds of judgments can be made about morality. And that suggests a different way to divide the jurisprudential landscape than has been traditional.

Here is one way that the new division might go. We might consider certain kinds of moral judgments in the law to be compatible with (some versions of) legal positivism for some purposes. We might say in particular that judgments about morality that involve assessments only of social facts—that require only descriptive, prescriptive, or persuasive evaluations—do not necessarily render legal interpretation compatible only with natural law. We might limit natural law, at least for some purposes, to the view that legal interpreters must make *substantive* moral judgments—judgments about what morality in fact entails, and not merely judgments about how others have assessed or will assess morality.[161]

However, we frame the point semantically, distinguishing descriptive or predictive judgments about morality from substantive moral judgments could prove useful, for example, in setting out the bounds of what Hercules can do. Recall that we have assumed that AI will soon be better at all descriptive and predictive analyses than human beings. That assumption could extend to purely descriptive and predictive claims about morality—about what people, as a matter of social fact, believe about morality and would be predicted to believe under particular circumstances.[162] But, we have assumed, AI does not and will not have a similar capacity to make substantive moral judgments—judgments about what is in fact moral or immoral. While we have assumed that Hercules can describe and predict moral judgments we have also assumed it cannot make substantive moral judgments.

---

161. Alternatively, we might put aside the traditional divisions in jurisprudence for these purposes and use new labels to distinguish these different kinds of judgments about morality as used in legal interpretation rather than redefining legal positivism and natural law. Given the massive literature on the topic, providing yet another set of definitions could sow confusion.

Yet another option would be to call purely descriptive or predictive claims about the law—including ones describing or predicting moral judgments—as "external" and claims about the law that involve substantive moral judgments as "internal." I am grateful to Bradley Wendel for suggesting this possibility.

162. Again, as discussed above, there may be limits on Hercules' ability to disentangle people's substantive moral judgments from other bases for their actions, including legal interpretations. *See supra* Part II.C.

There is an echo of this distinction in the writings of Lon Fuller. During his famous debate with H.L.A. Hart, Fuller addressed the issue of morality and, in particular, the moral beliefs prescribed by the Catholic Church. He appeared to be uncomfortable with the notion that a judge's religious beliefs—and, perhaps in particular, religious beliefs deriving from Catholicism—could inform legal interpretation. That can help to explain why he emphasized certain kinds of moral judgments he thought permissible in legal interpretation.[163] In addressing this issue, he suggested that the pronouncements of the Pope—in particular about divorce—were a kind of law rather than a kind of morality.[164] Real moral judgments, he implied, might be best understood as the "generally shared views of right conduct that have grown spontaneously through experience and discussion."[165]

Whatever the merits and precise content of Fuller's position on this issue, his analysis does arguably contain the seed of the idea discussed above. At least for some purposes, we might treat certain kinds of judgments about morality—descriptive and predictive claims about what people believe is moral—as compatible with legal positivism. A jurisprudent's views might then qualify as legal positivist if she defined the law as "prophecies of what courts will do in fact, and nothing more pretentious," even if those prophecies entailed predictions about the moral values judges are likely to hold and about how they will inform legal interpretation. Similarly, taking this approach could allow us to characterize a sociologist as acting as a legal positivist, if the sociologist makes purely descriptive and predictive claims about the laws of a society, even if those claims included descriptions or predictions about how the moral judgments of members of the society inform legal interpretations.[166]

163. Fuller, *supra* note 106, at 638 ("[I]n the thinking of many there is one question that predominates in any discussion of the relation of law and morals, to the point of coloring everything that is said or heard on the subject. I refer to the kind of question raised by the Pope's pronouncement concerning the duty of Catholic judges in divorce actions.").

164. *Id.*

165. *Id.*

166. The relationship between this approach and traditional definitions of legal positivism is interesting. Consider the social facts (or social) thesis, a common mechanism for defining legal positivism. Davis, *Legality*, *supra* note 5, at 61-62. One way to state the Social Facts Thesis is as holding that the content of the law depends ultimately only on social facts, not on moral facts. *Id.* Legal positivism can then be

In the context of AI, this alternative way of distinguishing legal positivism from natural law could have another benefit. We might say that natural law is "natural" in a way that was historically irrelevant: it can demarcate where legal interpretation should be performed by natural as opposed to artificial intelligence—when we, as human beings, must take responsibility for interpreting our laws. If legal interpretation requires only a description of morality as a social fact, or predictions of the moral judgments people will make, AI may soon be able to make the relevant moral judgments better than we can. Only if legal interpretation requires *substantive* moral judgments—according to our assumptions—is it "natural" in the sense of being beyond the ken of artificial intelligence. Artificial intelligence may not be able to interpret natural law but only what one might call "unnatural law."

IV. CONCLUSION

In a typical science fiction film, everything goes wrong when the computers we build to improve our lives begin making their own decisions. Think of the movie, "The Terminator." There, an artificial intelligence defense network becomes self-aware, acquires the capacity to act independently, and initiates a nuclear holocaust.[167] The prospect is not quite as apocalyptic of computer programs interpreting our laws for us, and perhaps taking over as our judges, awarding damages and even imposing criminal sanctions. But it is scary nonetheless. Why precisely can't computers interpret the law better than can human beings? If legal interpretation is merely a form of computation—as at least one highly knowledgeable and thoughtful futurist has suggested[168]—then computers are soon likely to be better at it than we are. Computers compute well. That is what they are designed to do.

---

understood as endorsing the Social Facts Thesis. But note that some claims about morality treat it merely as a matter of social fact. A purely descriptive claim about what a society—or a member of a society—in fact believes about morality is a matter of social fact, not a moral fact. We can make a descriptive or predictive claim about morality without taking any position on the actual content of morality. Only substantive moral judgments may require more than judgments about social facts.

167. Rebecca Hawkes, *The Terminator Timeline: A Guide for the (Understandably) Confused*, TELEGRAPH (Sept. 28, 2017, 12:57 PM), https://www.telegraph.co.uk/films/0/terminator-timeline-guide-extremely-confusing-story-far/.

168. TEGMARK, *supra* note 1, at 105.

Why not let them do it for us? Why not let the law operate without a role for human minds?

This Article suggests one possible reason not to do so. Legal interpretation at times may require a selection of ultimate ends—and the exercise of substantive moral judgment in selecting them. And computers may not be able to choose ultimate ends, even if they will soon outstrip human beings in achieving the ultimate ends they are tasked with pursuing.

To date, as stunning as technological developments have been, they seem to run along the lines of description and prediction—of identifying patterns in service of prescribed goals—but not of identifying goals worth pursuing. A change in the *kind* of analyses computers can perform—not just in the *degree* of difficulty of the tasks they can perform—may be necessary for them to make substantive moral judgments.

So, resolving a longstanding jurisprudential debate may soon have great practical significance. The debate turns on the morality's role in saying what the law is. Legal Dualism suggests a potential resolution to that debate. It holds that natural law provides the best account of law's nature when a legal interpreter seeks moral guidance from the law and that legal positivism provides the best account when a legal interpreter seeks merely to describe the law or to predict how others will interpret it. If Legal Dualism is right—to put the point somewhat crudely—it may allow us to identify an outer boundary on the role AI can play in legal interpretation. Only human beings may be able to interpret the law when it informs what we morally should and should not do. Jurisprudence may help to shape how society deals with technology. It may suggest that human beings should play a role in assessing the relevant law, for example, in programming self-driving cars.

Technology may also advance our understanding of jurisprudence. Focusing on the prospect of computers serving as judges may help sharpen our intuitions about the myriad judgments we think legal interpretation requires and the settings in which it requires them. Soon computers may be better than people at the purely descriptive and predictive (or persuasive) aspects of legal interpretation. That thought may strike some readers as a plausible prognostication, even if it is both startling and unsettling. Others may be more skeptical about what AI will be able to do. Whether prediction or thought experiment, however,

we may gain valuable insights by plumbing the intuitions that cause us to resist the notion that Hercules—a computer program that can make purely descriptive and predictive legal judgments for us and do so better than us—should serve as lawyer or judge. That exercise may help teach us something about the nature of law. As we consider the kinds of judgments that human beings can make and the kinds that computers cannot, we may revise how we distinguish legal positivism from natural law, at least for some purposes. We may also find a new reason to embrace Legal Dualism as a solution to the dispute that has preoccupied jurisprudents for more than a half-century: morality's role in legal interpretation. In at least this way, technological advances may benefit us.