Articles

Of Robolawyers and Robojudges

JOSHUA P. DAVIS†

Artificial intelligence (AI) may someday play various roles in litigation, particularly complex litigation. It may be able to provide strategic advice, advocate through legal briefs and in court, help judges assess class action settlements, and propose or impose compromises. It may even write judicial opinions and decide cases. For it to perform those litigation tasks, however, would require two breakthroughs: one involving a form of instrumental reasoning that we might loosely call common sense or more precisely call abduction and the other involving a form of reasoning that we will label purposive, that is, the formation of ends or objectives. This Article predicts that AI will likely make strides at abductive reasoning but not at purposive reasoning. If those predictions prove accurate, it contends, AI will be able to perform sophisticated tasks usually reserved for lawyers, but it should not be trusted to perform similar tasks reserved for judges. In short, we might welcome a role for robolawyers but resist the rise of robojudges.

† Research Professor, U.C. Hastings College of the Law, and Shareholder, Berger Montague PC. Many of the issues addressed in this Article will be explored more fully in Joshua P. Davis, *Unnatural Law: AI, Consciousness, Ethics, and Legal Theory* (forthcoming Cambridge University Press 2022/23). I am thankful for comments relevant to this draft from the participants at "The Internet and the Law: Legal Challenges in the New Digital Age," a symposium organized by the Center for Litigation and Courts at U.C. Hastings College of the Law and the Pound Civil Justice Institute, and from Andrew Bradt, Scott Dodson, David Engstrom, Clark Freshman, Deborah Hensler, Rick Marcus, Bradford Newman, Zach Price, Aaron Rappaport, Josh Rosenberg, Harry Surden, and Eugene Volokh.

TABLE OF CONTENTS

INTRODUCTION

Use of artificial intelligence (AI) has exploded in recent years. Its role in society is only likely to increase.[1] It has even begun to find its place in legal decision-making.[2] Indeed, scholars and other commentators have speculated that computers in the not-too-distant future may displace lawyers and judges.[3] AI may perform legal functions that traditionally have been reserved for us. This Article explores that possibility, addressing some issues that arise in particular in complex litigation.

In doing so, this Article focuses on two breakthroughs that would likely be necessary for robolawyers and robojudges to be effective. The first would be for AI to acquire what might informally be called common sense and what more formally may be captured by the terms abductive reasoning, inference to the best explanation, and the like.[4] A capacity for abductive reasoning might greatly expand the roles that AI can play in litigation, including (a) strategic advice—providing strategic advice to litigants and parties; (b) advocacy—advocating before a judge or jury; (c) class action settlement assessments—assisting a judge in deciding whether to approve class action settlements; and (d) proposing or imposing compromise—assessing the average result of a legal action, potentially as a benchmark for settlement or as a standard for imposing a result in arbitration. This Article identifies potential challenges for programming abductive AI and speculates that we may overcome them.

The second breakthrough would be for AI to choose its own objectives. At present, unless we provide AI ends to pursue, it is inert.[5] This Article argues that judges in resolving cases regularly need to engage in purposive reasoning, making moral or other value judgments. That provides reason to doubt that AI will be able to fulfill the judicial role without the capacity to make value judgments. The Article suggests that AI is unlikely to develop that capacity unless and until we imbue it with a first-person perspective—its own conscious experiences—something we have no idea how to do.[6] So we have reason to resist the rise of robojudges.

---

1. *See generally* MARTIN FORD, RULE OF THE ROBOTS: HOW ARTIFICIAL INTELLIGENCE WILL TRANSFORM EVERYTHING (2021).

2. *See generally* ETHAN KATSH & ORNA RABINOVICH-EINY, DIGITAL JUSTICE: TECHNOLOGY AND THE INTERNET OF DISPUTES (2017) (exploring how technology can construct new approaches to dispute prevention and resolution); BENJAMIN H. BARTON & STEPHANOS BIBAS, REBOOTING JUSTICE: MORE TECHNOLOGY, FEWER LAWYERS, AND THE FUTURE OF LAW (2017) (discussing using technology and procedural innovation to simplify and change the legal process).

3. *See, e.g.*, Eugene Volokh, *Chief Justice Robots*, 68 DUKE L.J. 1135, 1142 (2019).

4. *See, e.g.*, ERIK J. LARSON, THE MYTH OF ARTIFICIAL INTELLIGENCE: WHY COMPUTERS CAN'T THINK THE WAY WE DO 4 (2021); DOUGLAS WALTON, ABDUCTIVE REASONING, at xiii (2004); PETER LIPTON, INFERENCE TO THE BEST EXPLANATION (2d ed. 2004) (1991).

5. *See, e.g.*, STUART RUSSELL, HUMAN COMPATIBLE: ARTIFICIAL INTELLIGENCE AND THE PROBLEM OF CONTROL 10–11 (2019).

6. *Id.* at 16.

I. DEFINITIONS, ACCOMPLISHMENTS, FAILURES

To start, we should define a couple of terms and provide a brief summary of where we are today. Consider "artificial intelligence." We will define it in a non-technical way to include all non-organic entities—which at present means computers—that can perform tasks that historically only we could do because of our cognitive capacities.[7] Examples include playing chess, debating, and holding conversations.

A second term that warrants definition is "consciousness." Unless specified otherwise, we will use the word "consciousness" to mean phenomenal consciousness. Thomas Nagel famously suggested that "an organism has conscious mental states if and only if there is something that it is like to *be* that organism—something it is like *for* the organism."[8] We will use Nagel's definition, as have so many others.[9]

So defined, consciousness involves first-person experiences, such as how it feels for us to smell rotting eggs or touch a worm. Each experience is directly available to us by introspection but not by other means.[10]

With those definitions in place, let us turn to what AI has achieved to date and what it may achieve in the future. AI's flashier accomplishments include beating the best human players in the world at chess and Go, prevailing over human champions at *Jeopardy!*, and competing remarkably well against a world class debater.[11] AI has also shown some potential for identifying the words people read and the images they see based on their brain waves.[12]

More practical is the use of AI in making decisions regarding which employees to hire and promote, when to intervene to protect at-risk children in potentially unsafe homes, what the terms of bail should be for criminal defendants in light of the odds of flight or recidivism, how to target advertisements to maximize their efficacy, and whether to take remedial actions

---

7. We might broaden the definition to include tasks that intelligent non-human animals can perform because of their cognitive capacities. Nothing essential to the analysis should turn on this definitional issue.

8. Thomas Nagel, *What Is It Like to Be a Bat?*, 83 PHIL. REV. 435, 436 (1974).

9. *See, e.g.*, ANIL SETH, BEING YOU: A NEW SCIENCE OF CONSCIOUSNESS 13, 13 n.* (2021) (describing Nagel's article as "legendary" as "one of the most influential in all philosophy of mind").

10. For thorough—and at times skeptical—discussion of introspection, see generally INTROSPECTION AND CONSCIOUSNESS (Declan Smithies & Daniel Stoljar eds., 2012).

11. *See Computer Chess*, WIKIPEDIA, http://en.wikipedia.org/wiki/Computer_chess (last visited July 1, 2022) (AI prevailing at chess); Cade Metz, *In A Huge Breakthrough, Google's AI Beats a Top Player at the Game of Go*, WIRED (Jan. 27, 2016, 1:00 PM), https://www.wired.com/2016/01/in-a-huge-breakthrough-googles-ai-beats-a-top-player-at-the-game-of-go/ (AI prevailing at Go, as well as chess, Scrabble, Othello, and *Jeopardy!*); Srinivasa Ramanujam, *A 'Human Debater' Who Beat a Robot*, THE HINDU (Feb. 20, 2019), https://www.thehindu.com/news/cities/chennai/how-a-human-debater-won-over-a-robot/article26318594.ece (AI debating).

12. *See* Chelsea Whyte, *Mind-Reading Device Uses AI to Turn Brainwaves into Audible Speech*, NEW SCIENTIST (Apr. 24, 2019), https://www.newscientist.com/article/2200683-mind-reading-device-uses-ai-to-turn-brainwaves-into-audible-speech (AI converts brainwaves into audible words); Matthew North, *AI Recreates Videos People Are Watching by Reading Their Minds*, NEW SCIENTIST (Nov. 26, 2019), https://www.newscientist.com/article/2224866-ai-recreates-videos-people-are-watching-by-reading-their-minds (AI identifies videos people watch from their brainwaves).

for possible cancer and other diseases based on medical images.[13] AI may soon drive our cars, do our shopping, and engage in routine daily communications for us.[14] It has already made substantial progress in each of those tasks and many others.

Some of AI's current and future accomplishments are and will be highly beneficial, even if they may come with costs. In looking for cancer, for example, AI can enable us to decrease false negatives—and delayed treatment—as well as false positives—and unnecessary surgeries.[15] We should welcome such progress, notwithstanding that AI may put some radiologists and other medical services providers out of work.

Similar points apply to legal practice. Relying on AI to do some legal work—undertaking document review, for example, or drafting contracts—can help us meet some of the desperate need for modestly priced legal services.[16] Again, that is an improvement worth pursuing, even if it eliminates some human jobs.

Such job losses should be taken seriously. Still, they would not be unique. Technological change has always displaced workers.[17] Few people today are paid to make buggy whips. Even if AI shrinks the human work force—as some

13. *See* Rebecca Heilweil, *Artificial Intelligence Will Help Determine If You Get Your Next Job*, VOX (Dec. 12, 2019, 8:00 AM), https://www.vox.com/recode/2019/12/12/20993665/artificial-intelligence-ai-job-screen (use of AI in employment decisions); Elizabeth Brico, *When Data Discriminates*, MEDIUM (Apr. 17, 2019), https://medium.com/the-ai-issue-weapons-of-reason/when-data-discriminates-4791f14c5906 (use of AI to safeguard against child abuse); Matt O'Brien & Dake Kang, *AI in the Court: When Algorithms Rule on Jail Time,* ASSOC. PRESS NEWS (Jan. 31, 2018), https://apnews.com/article/ap-top-news-david-magee-courts-cleveland-us-news-20efb1d707c24bf2b169584cf75c8e6a (use of AI to set bail); Mike Kaput, *AI in Advertising: Everything You Need to Know*, MKTG ARTIFICIAL INTELLIGENCE INST. (Dec. 10, 2021), https://www.marketingaiinstitute.com/blog/ai-in-advertising (use of AI in advertising); *Unleashing the Promise of Artificial Intelligence in Radiology*, GEN. ELEC. (Sept. 2, 2021), https://www.gehealthcare.com/article/unleashing-the-promise-of-artificial-intelligence-in-radiology (use of AI in radiology).

14. Ben Lutkevich, *Self-Driving Car (Autonomous Car or Driverless Car)*, TECHTARGET (Oct. 2019), https://www.techtarget.com/searchenterpriseai/definition/driverless-car (use of AI in self-driving cars); Sarah Lewis, *Delivery Drone*, TECHTARGET, https://whatis.techtarget.com/definition/delivery-drone (last updated Dec. 2018) (use of AI in shopping); *Virtual Assistant*, WIKIPEDIA, http://en.wikipedia.org/wiki/Virtual_assistant (last visited July 1, 2022) (use of AI in virtual assistants).

15. *Artificial Intelligence Tool Improves Accuracy of Breast Cancer Imaging*, NYU LANGONE HEALTH (Sept. 24, 2021), https://nyulangone.org/news/artificial-intelligence-tool-improves-accuracy-breast-cancer-imaging#:~:text=When%20tested%20separately%20on%2044%2C755,suspect%20tumors%20by%2027%20percent (benefits of AI in detecting breast cancer).

16. *Managed Document Review Services Powered by Artificial Intelligence*, LEXCHECK, https://www.lexcheck.com/resources/managed-document-review-services-powered-by-artificial-intelligence-lc (last visited July 1, 2022) (use of AI in document review); Mikkel Boris, *How Long Before Machines Can Write Your Contracts?*, CONTRACTBOOK (Feb. 23, 2022), https://contractbook.com/blog/how-long-before-machines-can-write-your-contracts (use of AI in drafting contracts). *See generally* RICHARD SUSSKIND, TOMORROW'S LAWYERS: AN INTRODUCTION TO YOUR FUTURE (2013) (discussing, inter alia, use of AI and other technologies to meet unmet needs for legal services).

17. Susan Lund, Richard N. Cooper & Peter Gumbel, *What Can History Teach Us About Technology and Jobs?*, MCKINSEY GLOB. INST. (Feb. 16, 2018), https://www.mckinsey.com/featured-insights/future-of-work/what-can-history-teach-us-about-technology-and-jobs (discussing history of effects of technology on employment).

predict[18]—there should be ways to distribute society's gains broadly, perhaps through a universal basic income. Addressing that issue, however, is beyond the scope of this Article.

We will be concerned, instead, with whether AI would do more harm than good in other ways. Consider the potential for racial and other biases. Excitement about AI has given way to concern that it embodies invidious discrimination, including in making employment decisions, assessing at-risk children, setting bail, and advertising.[19] In some cases, it is a cruel irony that AI may exacerbate the very forms of discrimination it is intended to ameliorate.[20]

Similarly, consider the role of AI in guiding Internet users. Stuart Russell, an eminent AI expert, explains that content-selection algorithms on social media are often "designed to maximize *click-through*, that is, the probability that the user clicks on presented items."[21] Presumably, the algorithms were expected to select Internet links to suit users. But that is not all they did. They molded user preferences, enabling more reliable predictions of which links the users would click.[22] Russell suggests that this phenomenon can explain why social media tend to direct users to Websites that foster extreme political views.[23] Political extremists have more predictable preferences—and clicks—than do moderates.[24]

With the above sketch of AI's potential and perils in mind, we should consider two of its limitations. The first involves instrumental reasoning. AI currently struggles in pursuing certain goals we assign it because it lacks what we might colloquially call common sense. We cannot ask it to prevent harmful discrimination or the inculcation of hate. At present, we would have to define those and similar goals in much more technical and precise terms. AI is incapable on its own of formulating hypotheses about the kinds of links it should avoid. The second limitation relates to what we will call purposive reasoning. Current AI cannot decide for itself what purposes to pursue. It cannot form its

18. For a balanced discussion of whether, on net, AI will decrease or increase employment and how new jobs will be distributed see, e.g., THE BRITISH ACADEMY, THE IMPACT OF ARTIFICIAL INTELLIGENCE ON WORK 5 (2018), https://www.thebritishacademy.ac.uk/documents/280/AI-and-work-evidence-synthesis.pdf.

19. Miranda Bogen, *All the Ways Hiring Algorithms Can Introduce Bias*, HARV. BUS. REV. (May 6, 2019), https://hbr.org/2019/05/all-the-ways-hiring-algorithms-can-introduce-bias (bias in use of AI in employment decisions); Brico, *supra* note 13 (bias in use of AI to safeguard against child abuse); Shruti Verma, *How Data Distorts: Artificial Intelligence and Cash Bail Reform*, COLUM. POL. REV. (Nov. 16, 2020), http://www.cpreview.org/blog/2020/11/how-data-distorts-artificial-intelligence-and-cash-bail-reform (bias in use of AI in bail decisions); *Racial Bias in Marketing Unwittingly Introduced by AI Algorithms,* DAVID MEERMAN SCOTT (June 18, 2020), https://www.davidmeermanscott.com/blog/racial-bias-in-marketing-ai-algorithms (bias in use of AI in marketing). *See generally* VIRGINIA EUBANKS, AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE AND PUNISH THE POOR (2018).

20. *See, e.g.*, EUBANKS, *supra* note 19; SAFIYA UMOJA NOBLE, ALGORITHMS OF OPPRESSION: HOW SEARCH ENGINES REINFORCE RACISM (2018).

21. RUSSELL, *supra* note 5, at 8.

22. *Id*. at 8–9.

23. *Id.*

24. *Id.*

own objectives. Unless so directed, AI would not on its own decide not to promote bias or hate. The next section places these points in historical context.

II. FOUR PHASES AND FOUR FORMS OF REASONING

We can understand the progression of AI in terms of four phases, each defined by a form of reasoning: deductive, inductive, abductive,[25] and purposive. Deductive, inductive, and abductive reasoning are instrumental. They provide means of achieving prescribed goals. Purposive reasoning involves forming objectives. Prognostication is notoriously hazardous, particularly about AI. That said, Part III suggests that AI may well improve significantly at instrumental reasoning but it is unlikely to become capable of purposive reasoning. Put simply, Part III predicts that AI will acquire some common sense, but that it will not choose its own ends. The rest of the Article then assumes the accuracy of those predictions.

A. DEDUCTIVE REASONING: EARLY AI

For many decades, AI employed primarily deductive reasoning.[26] We created algorithms directing it how to respond to input. AI would then apply those algorithms in a mechanical fashion. Some of what deductive AI could accomplish was and is impressive. It can solve mathematical problems that would take most of us a great deal of time and effort, if we could handle them at all.

One of the simplest forms of deduction involves a syllogism. To illustrate the different forms of reasoning we will be discussing, it will be helpful to use variations on a concrete example. Here is one for deduction:

Example 1: Deductive Reasoning

Proposition 1: All fish swim.

Proposition 2: Tuna is a fish.

Conclusion: Tuna swims.

An advantage of deduction is that it can be unerring.[27] If a deductive argument is valid and sound, the conclusion that follows is correct.[28] If all fish swim, and if tuna is a fish, then tuna necessarily swims. To be sure, all dead fish and some injured fish do not swim. Also, for example, the red-lipped batfish found in the waters off the Galapagos Islands uses its pectoral fins to walk or, rather, to stagger like a drunk (and it is shaped like a bat and has bloated red lips

25. *See* LARSON, *supra* note 4, at 87–190 (discussing deduction, induction, and abduction).

26. *See* MELANIE MITCHELL, ARTIFICIAL INTELLIGENCE: A GUIDE FOR THINKING HUMANS 23–24 (2019); William Littlefield II, *A Type of Reasoning AI Can't Replace*, MIND MATTERS (Oct. 10, 2019), https://mindmatters.ai/2019/10/a-type-of-reasoning-ai-cant-replace/.

27. LARSON, *supra* note 4, at 107–10.

28. *Id*. at 110.

and a unicorn horn to boot!).[29] Still, the vast majority of fish swim. So the above syllogism generally holds true, even though it is imperfect.

Even if deductive AI can be reliable, it is limited in important ways. One of them is that it merely builds on or extends what we already know.[30] We have to supply the premises, such as that all fish swim. Deduction does not allow us to infer that all fish swim, unless we can derive that conclusion from more general rules.

Still, deductive AI enables some interesting and valuable applications. They include computers capable of playing chess or Go better than many of us can. For years, advances in the quality of AI chess resulted in part from improvements in algorithms we created and the speed and memory of computers.[31] We might program a computer to understand the rules of chess. We then might instruct it to give different weights to different chess pieces—say, a queen ten points and a pawn one point. The computer could then assign points to various available moves by analyzing all of the possible responses to them, responses to the responses, etc.

There are too many permutations in chess—or Go—for deductive AI to assess them all. So the inquiry would be limited by time and memory. Still, deductive AI made substantial progress in this way. But it never beat the best human beings in the world at chess.[32] That had to await development of *inductive* AI.

We can see important parallels in legal work. Word-processing programs—like Microsoft Word—relied historically on deductive reasoning.[33] We press a key and our computer applies a rule for what actions to take. That technology has had a profound impact on legal practice.[34] Large numbers of typists lost their jobs. Many lawyers today do their own word processing.

Other legal tasks, even seemingly menial ones, are not susceptible to deductive reasoning. Consider document review. It is hard to anticipate the characteristics of a document that would make it important to a case. General rules are clumsy for such purposes. Deductive AI thus had only limited utility at reviewing documents in place of (junior) attorneys.[35] That changed with improvements in inductive AI.

---

29. *Red-Lipped Batfish*, WIKIPEDIA, http://en.wikipedia.org/wiki/Red-lipped_batfish (last visited July 1, 2022).

30. LARSON, *supra* note 4, at 111–12.

31. *See Computer Chess*, *supra* note 11.

32. *Id.*

33. *See* Alison DeNisco Rayome, *Microsoft Boosts Office Productivity with AI for Word and Other Features*, TECHREPUBLIC (NOV. 30, 2018, 6:05 AM), https://www.techrepublic.com/article/microsoft-boosts-office-productivity-with-ai-for-word-and-other-features/ (discussing relatively recent development of use of AI in Microsoft Word).

34. *See* Nicole Black, *The Benefits of 21st-Century Word Processing Tools for Lawyers*, ABA J. (May 23, 2019, 6:00 AM), https://www.abajournal.com/news/article/21st-century-word-processing-tools-for-lawyers.

35. For discussions of uses of technology in civil procedure see, e.g., Richard Marcus, *Technology and Litigation: The 21st Century American Experience*, 25 ZIETSCHRIFT FUR ZIVILPROZES INT'L 99, 102–04, 109–10 (2021); David Freeman Engstrom & Jonah Gelbach, *Legal Tech, Civil Procedure, and the Future of*

B. INDUCTIVE REASONING: CURRENT AI

Recent decades, including the era of Big Data, have seen the rise of inductive AI.[36] Induction involves learning from experiences or observations. Over time, we discern a pattern and expect it to repeat in the future. We have developed formal rules to test for such patterns, which we call statistics (or econometrics). Inductive AI could be described as statistics on steroids.[37]

Inductive AI has made extraordinary strides for several reasons. We have become much more sophisticated at programming AI to perform statistical analyses.[38] The speed and memory of computers have increased dramatically. So has access to massive amounts of data, particularly from the Internet.[39] That last factor is why we use the term Big Data, but we might do better to call it the age of Inductive AI instead.

Regardless of labels, inductive AI performs many impressive tasks. Consider a simple example. We could train inductive AI by providing it with a large number of observations of objects and indicate which ones are swimming. It could then identify patterns associated with swimming objects. We could apply the inductive AI to videos of fish. It might conclude that all fish swim (or that a very high percentage of them do). If we inform the inductive AI that tuna is a fish (or similarly teach it to recognize tuna as a fish), it could then combine inductive reasoning with deductive reasoning to perform the following analysis:

Example 2: Inductive Reasoning

Proposition 1: The fish we observe (virtually) all swim.

Proposition 2: Tuna is a fish.

Conclusion: Tuna (almost certainly) swims.[40]

An advantage of inductive AI is that it can form its own rules, identifying patterns of which we are not aware. Examples include controversial algorithms that analyze images on dating sites and predict sexual orientation with a high rate of success. The programmers do not guide the AI. It finds its own markers based on subjects' self-identification.[41]

Inductive AI has various limitations. One of them is that it offers probabilistic predictions rather than certainty. Induction is not unerring as

---

Adversarialism, 169 U. PA. L. REV. 1001, 1004 (2021); Richard Marcus, *The Impact of Computers on the Legal Profession: Evolution or Revolution?*, 102 NW. U. L. REV. 1827, 1829 (2008).

36. *See* MITCHELL, *supra* note 26, at 24–34. *See generally* AJAY AGRAWAL, JOSHUA GANS & AVI GOLDFARB, PREDICTION MACHINES: THE SIMPLE ECONOMICS OF ARTIFICIAL INTELLIGENCE (2018).

37. *See generally* AGRAWAL ET AL., *supra* note 36.

38. *See generally* MITCHELL, *supra* note 26.

39. *Id.* at 98–100.

40. To be clear, AI would derive the major premise and possibly the minor premise of the syllogism through inductive AI, although syllogistic reasoning itself is deductive.

41. *See* Sam Levin, *New AI Can Guess Whether You're Gay or Straight from A Photograph*, THE GUARDIAN (Sept. 7, 2017), https://www.theguardian.com/technology/2017/sep/07/new-artificial-intelligence-can-tell-whether-youre-gay-or-straight-from-a-photograph.

deductive reasoning can be. At best, induction offers reliable statements about the likelihood of different possible outcomes.

Further, although inductive AI is, in the above sense, not as dependent on human input as deductive AI, we still need to play a significant role in programming it. We have to identify the variables inductive AI will use for analysis—such as pixels in videos—although it may then group those pixels based on patterns it discovers. We also have to label the data to match the variables. And we have to build the framework for the statistical analysis. We put in place what are sometimes called its "hyperparameters."[42]

Inductive AI also has a drawback: it may embody biases or draw inferences in ways we find inappropriate or offensive. To build on the example above, if inductive AI can predict sexual orientation with accuracy, it may then use that knowledge to discern other patterns and make predictions. They could be objectionable. They could also reflect outcomes that are shaped by societal prejudices.

Relatedly, inductive AI would not have the common sense to worry about predictions based on human biases or to report them as part of its results (unless we instruct it do so). We thus may not know that AI has discerned and relies on patterns based on sexual orientation, race, sex, religion, or other similar categories. Nor would AI be capable of making value judgments that might steer it away from stereotypes.

Consider the problems that can arise if inductive AI uses tainted data. It may use raises and promotions, for example, to assess worker potential. Those raises and promotions may reflect sexism, racism, homophobia, or religious intolerance. If so, inductive AI will make predictions that embody those biases, possibly amplifying invidious discrimination. As statisticians like to say: garbage in, garbage out.

Subject to the above limitations and concerns, inductive AI has accomplished a great deal. It is responsible for many of the remarkable successes in chess, Go, *Jeopardy!*, and debate, as well as in spotting cancer in medical images, interpreting brain waves, assisting employment decisions, identifying at-risk children in homes, predicting recidivism, and targeting advertisements.[43] However, it also poses the dangers we have recognized, including exacerbating stereotypes and inequalities.[44]

Inductive AI also has been responsible for significant improvements in automating litigation.[45] Think again about document review. In Part II.A, we noted that deductive AI has a limited capacity to review documents for relevance. We struggle to fashion general rules for deductive AI. Inductive AI can be much more effective. We can provide inductive AI a set of sample

42. *See* MITCHELL, *supra* note 26, at 97–98.
43. *See generally id.*
44. *Id.* at 106–08.
45. BARTON & BIBAS, *supra* note 2, at 131–32.

documents, some of which we label as "relevant" and others as "not relevant." AI can detect patterns for discerning relevance that we would miss. After iterations of training, inductive AI can then be set loose on a huge volume of documents, culling relevant documents with great speed, efficiency, and reliability.[46]

But note the active role we have to play in the process. Inductive AI is not capable on its own of making reasonable *preliminary* or *working* judgments about which documents are relevant. It lacks the common sense or reasonable judgment to perform such tasks. As discussed next, they would seem to require abduction.

C. ABDUCTIVE REASONING: THE NEXT BREAKTHROUGH?

The next category of reasoning is more difficult to define in part because we do not understand it well. A formal word for it dates back to the American pragmatist philosopher Charles Sanders Peirce: abduction.[47] We have already described it as common sense, although that term is loose. We might more precisely say that abduction involves two categories of hypothesis formation.

The first category comprises the formation of *testable* hypotheses. That was Peirce's focus. In discussing abduction, he was concerned with how we initiate the scientific method by generating hypotheses worth testing.[48] To take a concrete example, we might see smoke coming from the windows of a house. That could lead us to form the testable hypothesis that the house is on fire. (The relationship to common sense should be clear.) We might then follow up with an immediate investigation, peeking inside to look for flames.

A second category of abduction comprises *working* hypotheses. We accept certain propositions as true unless and until we have reason to doubt them. If we see black smoke billowing from a house, we may not wait to investigate before calling the fire department. We might feel confident enough to act right away (again, common sense). Our working hypothesis is that the house is burning.

We will use the term abduction to include forming both testable and working hypotheses. Some philosophers consider abduction to be a form of what they call "inference to the best explanation."[49] That entails, as the name suggests, drawing the inference that would provide the most compelling explanation for a state of affairs. To be clear, what counts as the "best" explanation of a phenomenon is far from straightforward, a complexity we will not explore.[50] Relatedly, abduction also may or may not be a form of inductive reasoning. But

46. *Id.* at 132.

47. *See* LARSON, *supra* note 4, at 25–26, 98–102, 160–68, 190; WALTON, *supra* note 4, at 3–17.

48. *See* Igor Douven, *Peirce on Abduction*, STANF. ENCYCLOPEDIA OF PHIL., https://plato.stanford.edu/entries/abduction/peirce.html#:~:text=The%20term%20%E2%80%9Cabduction%E2%80%9D%20was%20coined,from%20Peirce's%20writings%20on%20abduction (last visited July 1, 2022).

49. For a discussion of the topic, see generally LIPTON, *supra* note 4.

50. *See id.* at 21–54.

if it is, abduction is one we have been in significant part unable to get AI to do on its own.

Before developing that point, however, let us discuss variations on the example we have used. They will help us to see abduction's potential and limitations:

Example 3: Abductive Reasoning

Proposition 1: Fish swim.

Proposition 2: Tuna swims.

Conclusion: Tuna is a fish.

Abduction can help us recognize that tuna is a fish from our knowledge that fish swim and tuna does too. That conclusion does not follow deductively. While it may be true (let us assume) that all fish swim, it is not true that everything that swims is a fish. So there seems to be an element of informal or unsystematic judgment in arriving at the hypothesis that tuna is a fish from the observation that it swims.

Consider another example:

Example 4: Abductive Reasoning

Proposition 1: Fish swim.

Proposition 2: You swim.

Conclusion: You are a fish.

As we see, abduction can easily go awry. It does not incorporate formal standards for reliability like the ones that we have developed for deduction and induction. Still, *testable* hypotheses play a key part in advancing knowledge. That is true even for propositions that may at first seem absurd—such as that electrons do not occupy a single position at any given time.

Further, intuitively plausible *working* hypotheses are often essential for timely and effective action. You may have no experience with house fires, but you still might do well to call for help right away if you see smoke billowing from someone's home.

Note that the difference between example 3 and 4 is obvious to us but not necessarily to AI. We have to *program* AI to detect absurdities. That has proven difficult.[51] As a result, we have a capacity to navigate our everyday lives that AI lacks. AI may beat us at chess and Go, but we are superior to it at walking down the street, making appropriate small talk, and generally making sense of our physical and cultural environments. Melanie Mitchell aptly encapsulates this point: for AI hard things are easy and easy things are hard.[52]

Yet it seems plausible that we will make progress on simulating human thought through abductive AI. Doing so may require us to improve our understanding of how our neurological systems are structured and the

51. LARSON, *supra* note 4, at 126–29.

52. MITCHELL, *supra* note 26, at 33.

assumptions or schemas we have in place to facilitate our learning. Once we have that knowledge, we may be able to program abductive AI using the same structures and assumptions or schemas as we do. If so, abductive AI may be able to engage in all sorts of instrumental tasks that at present only we can do. One of them could be making preliminary or working judgments about which documents are likely to be relevant to a case. That could allow abductive AI to carry more of the burden of document review than can inductive AI. Abductive AI might not need as active and extensive human supervision.

Abductive AI also might be able to make the common-sense judgments necessary for effective advocacy, particularly at oral argument. It might formulate effective working hypotheses about the kinds of presentations that judges or jurors would find persuasive and credible. Even in the absence of a systematic empirical analysis, abductive AI might form hunches and make educated guesses. Those could prove crucial, especially in the real-world environment of a courtroom.

Note, however, that at times abductive reasoning—or common sense or inference to the best explanation—may involve more than just instrumental reasoning. It may require judgments about purposes or goals.

Here we come to a central debate in the philosophy of science. Hilary Putnam, for example, argued that science depends on "epistemic values,"[53] that is, values we pursue to help us make sense of the world. They include coherence, plausibility, reasonableness, simplicity, elegance, and beauty.[54] Those values may play an essential role in our selection of testable and working hypotheses, including what we call hunches and educated guesses. Our skill at abductive reasoning may rely in part on our assessments of whether a theory is elegant or beautiful. Yet we may not be able to program AI to make those assessments. Doing so may require a choice among objectives—or giving content to selected objectives, which may amount to the same thing. As we discuss next, AI may be incapable of that task.

D. PURPOSIVE REASONING: FUTURE OR FANTASY?

AI's greatest challenge involves purposive reasoning. That is the term we will use for the capacity to choose objectives. Unlike deduction, induction, and arguably some aspects of abduction, programmers have made no progress in getting AI to select its own ends. Stuart Russell, one of the world's leading experts on designing AI,[55] explains, "Because machines, unlike humans, have no objectives of their own, we give them objectives to achieve. In other words,

53. HILARY PUTNAM, THE COLLAPSE OF THE FACT/VALUE DICHOTOMY AND OTHER ESSAYS 30–34,132, 135, 143 (2002).

54. *Id.* at 132, 135, 141. For a fascinating discussion of the role of beauty in science in general and physics in particular, see STEVEN WEINBERG, DREAMS OF A FINAL THEORY: THE SCIENTIST'S SEARCH FOR THE ULTIMATE LAW OF NATURE 90–165 (1992).

55. *See* STUART RUSSELL & PETER NORVIG, ARTIFICIAL INTELLIGENCE: A MODERN APPROACH (4th ed. 2021).

we build optimizing machines, we feed objectives into them, and off they go."[56] We need to give AI its objectives. It cannot form them on its own. For now, that is a brute fact.

The same seems to be true for consciousness. As far as we know, AI does not have any.[57] Nor do programmers have any idea how to imbue AI with consciousness.[58] Russell again: "In the area of *consciousness*, we really do know nothing. No one in AI is working on making machines conscious, nor would anyone know where to start, and no behavior has consciousness as a prerequisite."[59]

These points suggest two important possibilities. The first is that there may be a relationship between AI's lack of consciousness and its inability to form objectives: AI's lack of consciousness may *explain* its inability to form objectives; AI's inability to form objectives may be important *evidence* of its lack of consciousness. The second possibility is that Russell may be wrong that consciousness is not a prerequisite for any behavior; consciousness may be essential for forming objectives.

Let us begin with why phenomenal consciousness might be necessary to form objectives. At least since the Scientific Revolution, we ascribe purpose only to conscious beings.[60] We have put aside Aristotelian and other forms of teleology as a way to describe mindless nature.[61] We no longer predict the effects of gravity in terms of the strivings of earth, water, air, and fire to sort themselves out in layers, from bottom to top in the order listed. Instead, we use the laws of Newtonian physics or, at extremes of scale, quantum mechanics or general relativity.[62] The laws of the physical sciences are causal, not purposive.

AI operates in the realm of science. It follows rules of causation. Those rules may be complicated. They are, however, ultimately deterministic—or, on the small scale of quantum, probabilistic. We set the train of AI in motion in the direction we choose. AI optimizes what we instruct it to optimize.

Human beings, in contrast, are motivated by desires, aversions, aspirations, and at least arguably values (more on that below). We alone have wishes. Stars and photons do not.[63]

---

56. RUSSELL, *supra* note 5, at 10.

57. *Id.* at 16.

58. *Id.*

59. *Id.*

60. JAMES LADYMAN, UNDERSTANDING PHILOSOPHY OF SCIENCE 14–15, 21 (2002).

61. *Id.* at 15–21.

62. *Id.* at 14–15, 67–68; FRANK WILCZEK, FUNDAMENTALS: TEN KEYS TO REALITY 93–125 (2021); WEINBERG, *supra* note 54, at 10–18.

63. To be sure, some schools of thought have attempted to eliminate the notions of mind and purpose in favor of causation. Notable were the radical behaviorists. Although the field is still alive, it has fallen far short of its original ambitions. Similarly, some philosophers claim that science will displace talk of the mental. *See generally* DANIEL DENNETT, CONSCIOUSNESS EXPLAINED (1991); NICHOLAS HUMPHREY, SOUL DUST: THE MAGIC OF CONSCIOUSNESS (2011). That possibility should be taken seriously. But for now, it seems, the best way to explain human behavior is in part in terms of motivations, that is, as purposive.

Let us consider an example to clarify what we mean by purposive reasoning.

Example 5: Purposive Reasoning

Proposition 1:We should not eat fish.

Proposition 2: Tuna is a fish.

Conclusion: We should not eat tuna.

Here Proposition 1 contains a value judgment, one that requires purposive reasoning.[64] The value judgment can be justified in various ways. Perhaps we should *not* eat fish because they are treated inhumanely, because it is wrong to destroy sentient life unnecessarily, or because doing so is bad for the planet. Those and other possibilities may be justified ultimately on various grounds, whether deontological, consequentialist, religious, or cultural.

Regardless, the point is that AI cannot make value judgments on its own, as Russell acknowledges.[65] Moreover, as Russell contends—arguably the main thesis of his book—we do not know how to program AI with sufficiently general, adaptable, and reliable objectives for us to trust it to operate independently.[66] His proposal is to build AI so that it seeks guidance from us about our preferences on an ongoing basis.[67] That is a possible strategy for contending with the risks posed by AI—one that should be taken seriously even if, as I contend elsewhere, the strategy has significant drawbacks.[68]

Particularly relevant for present purposes, Russell's ultimate position—that we should force AI to consult with us regularly—is in tension with his claim that consciousness is not a prerequisite for *any* behavior.[69] Consciousness may be a prerequisite for *forming objectives*. That behavior—as Russell notes—is essential.[70] Hence his belief in the need for human beings to steer AI.[71]

We have assumed that we will figure out how to build abductive AI. We will *not* make the same assumption about purposive AI. As noted above, Russell acknowledges that technologists do not have any idea how to imbue AI with consciousness—nor any idea where to start. As he reports, no one is even working on that project. So we will assume that AI will continue to lack consciousness and, as a result, it will remain incapable of purposive reasoning. We turn next to the implications of these assumptions for the future of AI in litigation.

---

64. Note that purposive reasoning is necessary to form Proposition 1—the major premise of the syllogism—but the syllogism itself involves deductive reasoning.

65. RUSSELL, *supra* note 5, at 10, 172.

66. *Id.* at 171–83.

67. *Id.* at 184–210.

68. *See* JOSHUA P. DAVIS, UNNATURAL LAW: AI, CONSCIOUSNESS, ETHICS, AND LEGAL THEORY (forthcoming Cambridge University Press 2022/23).

69. RUSSELL, *supra* note 5, at 16.

70. *Id.* at 171–83.

71. *Id.*

III. APPLICATIONS TO (COMPLEX) LITIGATION: PREDICTION AND MANIPULATION

If AI continues to improve its instrumental reasoning, it may become a powerful legal tool. We should expect AI to make accurate predictions about litigation, identifying potential outcomes and their odds of occurring. After all, it can detect patterns. That is why inductive AI is like statistics on steroids. In effect, it performs facial recognition by predicting which combinations of pixels in an image will be associated with a specific person, it reads minds by predicting which brain waves will be associated with which words or images, and it prevails at chess and Go by predicting which moves will have the highest probability of winning a game. As Agrawal, Gans, and Goldfarb point out, inductive AI's core strength is as a prediction machine.[72]

Abductive reasoning should fortify that strength. It could enable AI to formulate *testable* hypotheses that we miss, just as inductive AI currently draws inferences that we overlook.[73] Abductive AI might also develop effective *working* hypotheses that help it anticipate how judges and juries would respond to the law, evidence, and other cues about the merits of a case.

Nor should we ignore that prediction can be tantamount to manipulation. If we can anticipate how human beings will respond to stimuli, we may be able to stimulate desired behaviors, including from judges and jurors.

As discussed next, AI that can predict and manipulate could have great utility. Tasks it might perform include providing strategic advice, advocating for clients, helping judges assess class action settlements, and proposing or imposing attractive compromise outcomes.

A. ROBOLAWYERS: STRATEGIC ADVICE

One use of AI could be to help litigants and lawyers act strategically. They could consult AI, for example, in deciding whether to settle a case and, if so, on what terms. AI could also guide attorney conduct in litigation—in electing which witnesses to call, what legal and factual arguments to make, and what evidence to introduce.

At present, attorneys often act on hunches at worst and on experience at best.[74] But experience without reliable feedback does not yield expertise.[75] Lawyers may think they know what persuades judges and juries. That does not mean that they do. Evidence also suggests that experienced practitioners in a highly-skilled field, buttressed by a strong professional culture, will tend to have an exaggerated sense of their capacity to make sound predictions.[76] So lawyers may be inaccurate and overconfident at predictions.

---

72. *See* AGRAWAL ET AL., *supra* note 36.
73. LARSON, *supra* note 4, at 157–90.
74. *See* SUSSKIND, *supra* note 16, at 23–28 (2013).
75. *See* DANIEL KAHNEMAN, THINKING, FAST AND SLOW 234–44 (2011).
76. *Id*. at 209–21.

AI may be able to provide a more reliable empirical basis for strategic decisions. It could conceivably obtain information from real and mock oral arguments and trials. Using that data, AI might process images, sounds, words, and movements, and associate them with successful and unsuccessful advocacy. Based on those empirical analyses, it might provide valuable guidance on various issues—from what arguments attorneys should make to what clothes and accessories attorneys and witnesses should wear.

AI also may be able to tailor those recommendations to particular judges and jurors. It might do so based on observable facts about them—their appearances, their occupations, the places they live, or their accents. AI also could potentially rely on a treasure trove of information about on their online activities. Litigators could potentially feed AI the data that is collected about what products and services individual decision-makers buy, what emails and texts they send, where they travel throughout the day as recorded by their cell phones, and anything else that is collected and packaged about us from the Internet and that AI determines is relevant. After all, we live in what has been called the Age of Surveillance Capitalism.[77] All of this data might give AI insights into how to manipulate judges and jurors in ways we would not predict and may not even be able to understand.

AI also should be better able to assess its own limitations than can human experts. Again, we tend to be more confident than accurate.[78] As Oliver Wendell Holmes, Jr. observed, "Certitude is not the test of certainty. We have been cocksure of many things that were not true."[79] AI should be better than we are at making predictions and at acknowledging uncertainty. Both can be valuable.

Note, however, that there are ethical limits on providing legal advice with which robolawyers may struggle. The ethical rules, for example, prohibit lawyers from knowingly assisting a client in a crime or fraud.[80] That can include giving guidance about how best to avoid prosecution, such as by destroying evidence.[81] Yet it is not clear that AI could draw the distinction between permissible advice and impermissible assistance (or that it can do anything knowingly). We will explore this issue in Part V as it relates to robojudges, but it can be relevant to robolawyers as well.[82]

77. *See generally* SHOSHANA ZUBOFF, THE AGE OF SURVEILLANCE CAPITALISM: THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER (2019).

78. KAHNEMAN, *supra* note 7536, at 209–21.

79. OLIVER WENDELL HOLMES, JR., COLLECTED LEGAL PAPERS 311 (1920).

80. *See* MODEL RULE OF PROFESSIONAL CONDUCT 1.2(d).

81. *See* W. BRADLEY WENDEL, LAWYERS AND FIDELITY TO LAW 189–94 (2010).

82. For a discussion of the relationship between ethics and jurisprudence for lawyering see, e.g., DAVID LUBAN, LEGAL ETHICS AND HUMAN DIGNITY 99–130 (discussing Lon Fuller's jurisprudential views and legal ethics).

B. ROBOLAWYERS: ADVOCACY

AI may be able to do more than just provide strategic advice; it may be able to advocate. AI may someday synthesize legal sources, evidence, allegations, or the like in a written form that maximizes the chances of a desired outcome. It may do so by appealing not only to a judge's values but also to her biases and confusions. In this regard, too, AI may exploit information about a judge, perhaps from public records, including past judicial proceedings, and perhaps from her online activities, much of which is or may be for sale.[83]

In the not too distant future, we also may be able to build AI that looks and sounds like a particularly credible advocate. And, again, the look and sound of a roboadvocate could even be bespoke—tailored to a particular judge or set of jurors.

To be sure, those sorts of achievements would not come easily. Technologists have struggled with natural language.[84] AI cannot reply to written or spoken words in some ways that human beings find easy. It is possible that doing so would require AI to move beyond syntax to semantics—not only identifying patterns in words but understanding what those words mean.[85]

On the other hand, it is also possible that AI will become so sophisticated, fast, and powerful that it will write more effective briefs than we can without understanding their contents. It may simply predict our responses to letters, words, sentences, or paragraphs, much as it associates pixels in an image with a person's identity. Let us assume that AI does—at least to a significant extent.

We still may retain an advantage over AI in advocacy. Although in theory attorneys do not testify when they advocate, seasoned observers suggest that credibility is essential to persuasion.[86] We have assumed AI is not conscious and so it cannot be sincere. That might undermine a judge's confidence in AI arguments.

That problem might not be as significant for written argument. It is not clear that a judge ever needs to know that a machine was the author of a brief, at least as long as an attorney signs it (and, presumably, reviews it before filing).[87] But it would be more difficult to mask that AI is arguing in court. So AI might operate at a net disadvantage in oral advocacy.

But, then again, it might not. We know that people form strong relationships with inanimate objects, even come to love them.[88] We have a powerful propensity to anthropomorphize. A robot might take advantage of that propensity. It might be designed so we find it credible. Such a robot could rely

---

83. The French have outlawed use of predictive analysis regarding magistrates and members of the judiciary. *See* French Justice Reform Act, Section 33.

84. *See* LARSON, *supra* note 4, at 50–59.

85. *See id*. at 204–34; MITCHELL, *supra* note 26, at 170–71.

86. *See* HERBERT J. STERN, TRYING CASES TO WIN 13, 15 (1991).

87. *See* FED. R. CIV. P. 11 (requiring signature of attorney who takes responsibility for filing).

88. *See, e.g.*, Joelle Renstrom, *Why Humans Love Robots Like People*, DAILY BEAST (May 8, 2019), https://www.thedailybeast.com/why-humans-love-robots-like-people.

on information about a decisionmaker's background, facial expressions, accent, and the like to adjust its gestures, posture, voice, and appearance. Presumably judges should not be influenced by these factors. Yet they likely are to be. So robolawyers may someday be more effective than human lawyers in all forms of advocacy.

As with advising, we should note the ethical limits on attorneys advocating. They are required to certify that they *believe* there is an appropriate basis in law and evidence for the positions they take in written filings and oral arguments.[89] We have assumed robolawyers will not have subjective beliefs. That too gives rise to difficulties that relate to robojudges, as discussed in Part V.

C. ROBO-OBJECTORS: SETTLEMENT APPROVAL IN CLASS ACTIONS

Another potential use of AI would be to assist judges in a difficult task in class actions. They have an obligation to protect absent class members by assessing whether a proposed class action settlement is fair, reasonable, and adequate.[90] That is no mean feat.

When class litigation turns to settlement, the adversarial system breaks down. In litigation, plaintiffs and defendants have incentive to offer competing views of the law and facts. That can assist judges in reaching informed conclusions. When the parties settle, however, they present a unified front. The judge is largely on her own in determining whether plaintiffs have obtained sufficient relief for the class.

True, objectors may challenge a class action settlement. But they often have a limited capacity to assess the relevant law and evidence—and a limited interest in doing so.[91] Indeed, they often seek merely to gum up the works until they are paid to go away. So, a judge may find herself in the unenviable position of second-guessing attorneys who know the law and the facts far better than she can.

Enter AI. A judge could use its analysis of the likely outcomes of litigation and their odds of occurring. That could assist a judge in gauging the relationship between a settlement and what might be expected to happen on average in litigation. AI could also identify any extreme results that might occur, a consideration relevant to whether a settlement reflects reasonable responses to risk. AI too could assess the confidence the judge should have in its predictions, information also suggestive of the range of plausible views about an appropriate settlement.

---

89. *See, e.g.*, FED. R. CIV. P. 11(b) (attorney when "signing, filing, submitting or *later advocating*" in court "certifies to the best of the person's knowledge, information *and belief*" that her arguments have an appropriate basis in law and evidence) (emphasis added).

90. FED. R. CIV. P. 23(e)(2).

91. *See* Robert H. Klonoff, *Class Action Objectors: The Good, the Bad, and the Ugly*, 89 FORDHAM L. REV. 475, 498 (2020).

To be clear, for AI to play this sort of role would likely require procedural innovation. A court, for example, might direct the parties to identify the most relevant legal precedents for evaluation by AI. If litigation has been pending for a while, that task might be easy or even unnecessary. The parties may have already cited the key case law. Or AI might be able to do legal research on its own.

Matters are more complicated when it comes to the facts. The court might need the parties to present the evidence in a form that AI can evaluate—likely documents, including electronically stored information, as well as transcripts of testimony, such as from depositions. Again, if the litigation has progressed that task might be unnecessary. Summary judgment briefing might suffice. But it might not. The parties at summary judgment may not have addressed key issues for trial, or settlement may have come before briefing on summary judgment.

A new procedure might be necessary by which parties feed evidence to AI. It is beyond the scope of this Article to analyze whether Federal Rule of Civil Procedure Rule 23 empowers judges to use AI in this way, although courts assessing class action settlements may request information from the parties.[92] The main point, however, is that AI might help judges exercise independent judgment in assessing class action settlements.

D. ROBO-ADR: EXPECTED VALUE MEDIATION OR ARBITRATION

AI could go beyond advising, advocating, and assisting. It could offer an alternative form of dispute resolution. We might call it expected value (EV) mediation or arbitration.

EV alternative dispute resolution would be most straightforward for monetary recoveries. It would involve AI calculating the expected value of the outcome of trial. To take a simplistic example, AI might determine that a plaintiff has a 50% chance of losing and a 50% chance of recovering $100,000. The expected value would then be 0.5 x $0 + 0.5 x $100,000 = $50,000. The parties might use that number as a *guide* for settlement in mediation. Alternatively, they might empower an arbitrator to *impose* the expected value of litigation to resolve a dispute, which we might call "Expected Value Arbitration" or "EVA."[93]

AI EVA might have numerous advantages over trial.[94] Those could include allowing parties to seek an independent judgment without the winner-take-all risks of resolution by a finder of fact,[95] minimizing harms from errors in legal decision-making,[96] and encouraging desirable expenditures on attorney's fees

92. *See* FED. R. CIV. P. 23(e)(1)–(2) (parties must provide court information sufficient for it to assess whether a proposed settlement is fair, reasonable, and adequate).

93. *See* Joshua P. Davis, *Expected Value Arbitration*, 57 OKLA. L. REV. 47, 47 (2004).

94. *Id*. at 70–106.

95. *Id*. at 71–85.

96. *Id*. at 85–94.

and costs, often at lower amounts than would traditional litigation.[97] Those benefits might be particularly great in class actions where there is a great deal at stake, the parties are likely to be averse to risk, errors may prove particularly costly, and litigation expenditures can be extraordinary.

That said, AI EVA would give rise to some thorny issues. For example, should it consider the relative resources or the quality of counsel of the parties? Presumably, more expensive attorneys tend to skew the results of litigation in favor of their clients as compared to less expensive attorneys. Otherwise, we would have to assume that parties act systematically irrationally in paying for legal services. But we may be troubled if AI EVA were to adjust its analysis in light of the attorney's fees that the parties would anticipate expending. That could reward those with wealth in a way that is difficult to justify.[98] It could add yet one more advantage to the many that the "haves" hold over the "have nots."[99]

As with lawyers advising and advocating, we see that value judgments may play an important role in ADR. So now let us turn to the context in which that issue arises most squarely: robojudges ruling in litigation.

IV. SOME LIMITS AND DANGERS OF ROBOJUDGES

The above discussion identifies some legal tasks that we might expect AI to perform relatively well. It is particularly likely to succeed if it improves greatly at abductive reasoning, even if it does not acquire purposive reasoning. Part IV turns to a task that could be beyond the capacity of non-purposive AI: judging.

Part IV.A explains why judges likely need to make value judgments to reach particular conclusions. Given our assumption that AI cannot make such judgments, AI would seem unable to fulfill the judicial role.

Part IV.B then addresses a potential alternative endorsed by Eugene Volokh.[100] We might ask AI to use its power of prediction to write *persuasive* judicial opinions. Might such a robojudge perform as well as or even better than human judges? Part IV.B offers reasons to doubt it would. It suggests that robojudges might be better than us at writing opinions that *seem* right but worse than us at writing opinions that *are* right. If so, robojudges might corrupt judicial decision-making rather than enhance it.

A. JUDGING AND PERVASIVE VALUE JUDGMENTS

We have assumed that AI will not be able to make value judgments. The next issue is whether judges make value judgments when ruling in cases. That issue may seem to depend in part on an enduring controversy—the role of moral

97. *Id*. at 94–106.

98. *Id*. at 119–21.

99. Marc Galanter, *Why the "Haves" Come out Ahead: Speculations on the Limits of Legal Change*, 9 LAW & SOC'Y 95, 103 (1974).

100. Volokh, *supra* note 3, at 1161.

judgments in saying what the law is. That has been the primary debate in jurisprudence for over half a century.[101]

Fortunately, we need not resolve that debate to conclude that value judgments likely play a pervasive role in judging. Most jurisprudents acknowledge that moral judgments play a significant role in *creating* and *applying* the law, even if they do not or should not play a regular role in saying what the law is—in *interpreting* the law.[102] Further, various legal values—including planning, authority, consistency, and predictability—may require judgments about ends, whether or not those values are moral.

1. Making Law

Consider the *creation* of new law. Some legal positivists deny that judges should make moral judgments in interpreting existing law.[103] But they accept that judges, legislators, and the like can make value judgments, including moral ones, when establishing *new* law.[104] On that point, there is a widespread consensus.[105]

That concession may seem relatively narrow. It is not. A reason is that there is no sharp distinction between making law and interpreting it.[106] That is in part because uncertainty in law is a matter of degree.[107] Generally speaking, judges elide whether they are, on one hand, extending existing law to fill gaps, resolve inconsistencies, and clarify ambiguities or, on the other hand, performing those same functions by creating new law.[108] And little usually turns on that difference.[109] Either way, judges overwhelmingly apply the law retroactively and speak as if they are simply discovering what that law is.[110]

2. Applying Law

A widespread consensus also exists that judges make moral or other value judgments in *applying* the law.[111] A value judgment—even a moral judgment—

101. *See, e.g.*, Scott Hershovitz, *The End of Jurisprudence*, 124 YALE L.J. 1160, 1162 (2015); Scott Shapiro, *The "Hart-Dworkin" Debate: A Short Guide for the Perplexed*, *in* RONALD DWORKIN 22 (Arthur Ripstein ed., 2007).

102. Joshua P. Davis, *Artificial Wisdom? A Potential Limit on AI in Law (and Elsewhere)*, 72 OKLA. L. REV. 51, 58–61 (2019) (discussing the prominent legal positivists H.LA. Hart, Scott Shapiro, and Joseph Raz).

103. *Id.*

104. *See, e.g.*, SCOTT J. SHAPIRO, LEGALITY 247–56 (2011); H.L.A. Hart, *Positivism and the Separation of Law and Morals*, 71 HARV. L. REV. 593, 608–15 (1958).

105. Davis (2019), *supra* note 102, at 58–61 (discussing Hart, Shapiro, and Raz).

106. Joshua P. Davis, *Legality, Morality, Duality*, 2014 UTAH L. REV. 55, 79–80 (2014).

107. *Id*. at 77.

108. *Id*. at 78.

109. *Id.*

110. SHAPIRO, *supra* note 104, at 270–71. Note there are some exceptions, including the qualified immunity doctrine. *See* Pearson v. Callahan, 555 U.S. 223 (2009). Courts also at times acknowledge they face legal issues of first impression.

111. Davis (2019), *supra* note 102, at 58–61 (discussing the prominent legal positivists H.LA. Hart, Scott Shapiro, and Joseph Raz).

may be embedded in a legal rule or standard.[112] Courts may refuse to enforce contracts if they are unconscionable.[113] Defendants may be liable in tort if they do not take reasonable care.[114] Plaintiffs may recover punitive damages if a defendant acted wantonly.[115] Assessing unconscionability, reasonable care, and wantonness involves value judgments, likely moral ones.[116]

To be sure, jurisprudents disagree about whether those moral judgments are part of the law. So-called exclusive (or hard) legal positivists take the position that moral judgments are never part of the law.[117] They might say that is true even if the law relies on them—just as mathematics presumably is not part of the law although the law at times uses mathematics. In contrast, inclusive (or soft) legal positivists hold the view that the law can contain moral judgments (but that whether it does is ultimately a matter of pure social fact).[118] One might reasonably suspect that the disagreement here is more semantic than substantive. Regardless, legal positivists tend to accept that *applying* the law need not be morality-free—much less value-free.[119]

Again, this point may seem narrow. It, too, is not. No clear distinction exists between, on one hand, creating or interpreting law and, on the other, applying it. Consider so-called mixed questions of fact and law. Courts sometimes say that a mixed question exists when historical or primary facts are established or undisputed, but ultimate inferences and legal consequences are contested.[120] The line between a historical or primary fact and an ultimate inference, however, is fuzzy. Similarly, applications can shape rules and vice-versa. Factual scenarios can accrete into rules and rules can dissolve into factual issues, such that value judgments relevant to one can inform the other.[121]

---

112. SHAPIRO, *supra* note 104, at 270.

113. *See Unconscionability*, LEGAL INFO. INST., https://www.law.cornell.edu/wex/unconscionability (last visited July 1, 2022).

114. *See Standard of Care*, LEGAL INFO. INST., https://www.law.cornell.edu/wex/standard_of_care#:~:text=Standard%20of%20care%20is%20an,will%20be%20liable%20for%20negligence (last visited July 1, 2022).

115. *See Punitive Damages*, LEGAL INFO. INST., https://www.law.cornell.edu/wex/punitive_damages (last visited July 1, 2022).

116. SHAPIRO, *supra* note 104, at 270–71. (discussing unconscionability and reasonable care).

117. *See, e.g.*, *id*. at 271–72, and JOSEPH RAZ, THE AUTHORITY OF LAW: ESSAYS ON LAW AND MORALITY 46 (1979).

118. *See, e.g.*, H.L.A. HART, THE CONCEPT OF LAW 238–76 (3d ed. 2012) (Postscript), and JULES COLEMAN, THE PRACTICE OF PRINCIPLE: IN DEFENCE OF A PRAGMATIST APPROACH TO LEGAL THEORY 67–69 (2001).

119. Davis (2019), *supra* note 102, at 58–60.

120. *See* Pullman-Standard v. Swint, 456 U.S. 273, 289 n.19 (1982) (mixed question of law and fact arises when the historical facts are established, the rule of law is undisputed, and the issue is whether the facts satisfy the legal rule); Khan v. Holder, 584 F.3d 773, 780 (9th Cir. 2009); Suzy's Zoo v. Commissioner, 273 F.3d 875, 878 (9th Cir. 2001) (mixed question "exists when primary facts are undisputed and ultimate inferences and legal consequences are in dispute"). Mixed questions of law and fact often require judgments about the values that animate legal principles. *See* Smith v. Commissioner, 300 F.3d 1023, 1028 (9th Cir. 2002).

121. In antitrust, for example, courts will arrive at a per se rule—holding that conduct automatically violates the law—after enough experience of finding the conduct consistently anticompetitive. N. Pac. Ry. Co. v. United States, 356 U.S. 1, 4–6 (1958). The process can also work in the opposite direction so that courts may require what is called application of the rule of reason—which entails consideration of procompetitive and anticompetitive effects—if additional judicial experience suggests the behavior is not consistently

3. Interpreting Law

Jurisprudents also tend to agree that value judgments can inform legal interpretation. H.L.A. Hart acknowledged that purposive reasoning can play an important role in saying what the law is, although he denied that the purposes of the law are necessarily moral.[122] He famously suggested, for example, that a key purpose of Nazi law was *evil*.[123] Joseph Raz argues that a distinctive understanding of legal *authority* forecloses moral judgments in legal interpretation.[124] Scott Shapiro contends that law is a *plan* (or plan-like norm) that provisionally resolves moral judgments and eliminates the need to revisit them in saying what the law is.[125] Evilness, authority, and planning are contestable and subtle values, even if not moral ones. Judges interpreting the law to serve those values would be expected to make judgments about them in legal interpretation.

To be sure, some readers may be skeptical about the legal positivism of Hart, Raz, and Shapiro. The law may not *always* serve moral purposes, but ideally it would *often* do so, at least in part. Further, whether a judge should consider morality in saying what the law is in any given setting would seem to depend in part on her judgments about political morality, including about the appropriate role for, say, an unelected judge in a representative democracy.

We should also note that authority and planning also seem like abstract moral values, as do internal consistency and predictability. Lon Fuller thus characterized such values as forming the internal morality of law.[126] Perhaps those values are not always moral. However, for legal interpreters attempting to fulfill their moral obligations—likely including many judges in representative democracies—they naturally would be interpreted as moral. If a judge, for example, attempts to abide by the moral responsibilities of her judicial office—assuming she has some—she would want to consider the moral force and nature of authority and planning. She thus would exercise moral judgment in deciding how she should implement those values and possibly how she should balance them against others, including achieving justice in particular cases.[127] From a

anticompetitive, all things considered. *See* Leegin Creative Leather Products, Inc. v. PSKS, Inc., 551 U.S. 877 (2007).

122. Hart (1958), *supra* note 104, at 613–15.

123. *Id*. at 616.

124. JOSEPH RAZ, ETHICS IN THE PUBLIC DOMAIN: ESSAYS IN THE MORTALITY OF LAW AND POLITICS 210–37 (rev'd ed. 1994).

125. *See generally* SHAPIRO, *supra* note 104.

126. *See, e.g.*, LON L. FULLER, THE MORALITY OF LAW 46 (rev'd ed. 1969).

127. Note in this regard Scalia's comment that he would prove a faint-hearted originalist if that approach would allow whipping prisoners, see Antonin Scalia, *Originalism: The Lesser Evil*, 57 U. CIN. L. REV. 849, 861–62 (1989)—a comment he later rescinded. *See* Jennifer Senior, *In Conversation: Antonin Scalia*, N.Y. MAG. (Oct. 4, 2013) (available at https://nymag.com/news/features/antonin-scalia-2013-10/). Further note that he was willing to base his interpretive analysis on contested judgments about predictability and consistency, important and pervasive values in the law, and possibly moral values as well. *See generally* Antonin Scalia, *Response*, *in* A MATTER OF INTERPRETATION (Antonin Scalia & Amy Gutmann, eds., 1998).

moral perspective, it is hard to see how those issues could be anything other than moral ones.[128]

In any case, moral or otherwise, authority and planning are values. The point is that although, for example, Justice Scalia endorsed different judicial value judgments than did Justice Cardozo—focusing more on consistency and predictability[129] than on societal changes and morality[130]—Scalia endorsed judicial value judgments nonetheless. As a result, AI cannot choose among them or fill in their content if it cannot engage in purposive reasoning.

B. MANIPULATION: WHAT SEEMS RIGHT, NOT WHAT IS RIGHT

There is a strong case, then, that judging often involves value judgments, including likely moral ones. The next issue is whether robojudges might nonetheless be more effective than human judges at deciding cases.

We do not have space to address that issue systematically. We can, however, consider an argument that Eugene Volokh makes in *Chief Justice Robots*.[131] Volokh contends that if AI is able to write more persuasive opinions than we can, we should accept AI as a judge.[132] His position is characteristically thoughtful and forceful.

In responding to Volokh, we will develop a distinction that has broad application: between what *is* right and what *seems* right. That opens up the possibility that robojudges may be inferior to human judges, even if robojudges write more persuasive opinions than we do.

Our analysis will rely on some plausible assumptions without defending them. The first one is that moral and other value judgments can be better and worse, maybe even right and wrong. The second is that moral and other value judgments matter—that we should act on better value judgments rather than worse ones. The third is that we have some capacity, however imperfect, to make accurate value judgments. The fourth is that we are capable of erring in making value judgments, including if we are misled by self-interest or other biases.

To be sure, all of these assumptions are controversial. Credible philosophers would contest each one. But without them we would have little prospect of determining what we morally should do or explaining why we should try to make sound value judgments at all. We might as well give up. And, in any case, we lack space to justify these assumptions.

---

128. For a discussion of the role of morality in judicial decisionmaking, see generally Davis (2014), *supra* note 106, at 55–61.

129. *See generally* Scalia (1998), *supra* note 127.

130. *See* BENJAMIN N. CARDOZO, THE NATURE OF THE JUDICIAL PROCESS 94–97, 133–34 (1921).

131. *See generally* Volokh, *supra* note 3.

132. *Id*.

1. Why We May Write Better Judicial Opinions than AI Does

Let us begin with why our judicial opinions may be better than AI's. We have assumed that we can make value judgments but AI cannot. We have also concluded that value judgments are likely pervasive in judicial decision-making. AI, then, has to rely on our value judgments in writing opinions. That places AI at a disadvantage. Its value judgments are derivative. Ours are primary. Perhaps, however, AI can describe and predict our value judgments, and thereby make derivative value judgments of a quality similar to or even better than ours. For several reasons AI is unlikely to succeed in that task.

Changing Circumstances. First, circumstances change. AI often cannot apply old value judgments in a mechanical way to new settings.[133] How old values apply to novel facts will not always be self-evident. The values may have to be clarified or refined. We can do that. AI cannot. It lacks the ability to form ends. As a result, AI will need new data from us to discern our views as the environment changes, whether those changes are, say, cultural, or, yes, technological.

Changing Values. Second, values change. Of course, they may not always change for the better. But we lack a viable alternative to relying on evolving values. Otherwise, we might have to accept that slavery, monarchy, and the like are as good as modern practices. We are the firsthand source of changes in values. AI can detect those changes only through us. That provides a second reason its value judgments will grow stale.

Noise. Third, data about human value judgments are noisy. Past judicial opinions and other sources of law are tainted by biases, psychological desires, and related products of motivated cognition. We have some hope of distilling our value judgments from such noise. With diligence, self-discipline, and candor, we may disentangle what we think is right from what we want to believe.

True, our efforts along these lines are likely to be flawed. But it is not clear how AI can distinguish value judgments from biases at all. Beliefs do not come pre-labeled. Only substantive value judgments enable us to distinguish one from the other. As we have noted, AI cannot make substantive value judgments. It has to reach conclusions about values based on what we say and do. Its inferences will thus be tainted to the extent our statements and actions are. We, in contrast, may be able to discern the signal of values from the noise of our biases, however imperfectly we do so.

2. Why AI Nevertheless May Write More Persuasive Judicial Opinions than We Do

Our capacity to make sound value judgments could lead us to write more persuasive judicial opinions than AI does. But it may not. The silver tongue of the devil may convince us more effectively than the candor of our better angels.

133. RUSSELL, *supra* note 5, at 171–83.

Put less poetically, AI may win us over by telling us what we want to hear rather than what we should hear. Some of our value judgments are uncomfortable. They can reveal that we have been acting inconsistently with what we consider best on reflection. Perhaps we have interpreted the law or applied it in ways that reflect systemic biases. Perhaps we have adopted flattering views about ourselves that would not survive scrutiny. We may resist those and similar possibilities and, as a result, prefer judicial opinions that deftly rationalize our mistakes to judicial opinions that in a deeper sense are right.

Moreover, it is often difficult to make clear statements about the law or its application when we are recalibrating. We do not always see how a new approach will play out. That is a reason courts at times emphasize that they decide one case at a time and that statements about circumstances not before them are dicta. It can take a while for legal change to cohere. Until it does, the opinion that is most persuasive—perhaps because it offers an orderly statement of the law—may not be the best one.

Further, if AI is directed to write judicial opinions that judges or others will find most persuasive, we should expect it to *exploit* weaknesses in how we think. AI will not do so out of some improper motivation. It has no motivations. To function, it will need data about what judges and others find persuasive. Those data would presumably embody all sorts of inclinations that judges and others may try to resist. A robojudge, however, will lack the capacity to distinguish sound arguments from manipulative or dangerous ones. It will, by its nature, discover and take advantage of ingenious ways to make illegitimate arguments seem legitimate. That could make AI judicial opinions more persuasive than ours even though—perhaps *because*—they are inferior.

3. Why AI May Corrupt Judicial Decision-Making

This last point suggests the possibility that robojudges could *corrupt* judicial decision-making. That could occur in various ways, each potentially compounding the others.

Staleness. Robojudges could deprive themselves of data. If they take over all or most of the judiciary, they will no longer have recent human judicial opinions from which to detect patterns. As our circumstances and values change, AI opinions will grow progressively staler. They may no longer reflect modern society and its beliefs.[134] Of course, we could potentially detect such staleness and compensate for it, maybe by infusing the body of AI opinions with human ones. The other sources of corruption discussed below, however, may interfere with our capacity or motivation to do so.

Misplaced Deference. We may confuse AI's acumen at persuasive legal analysis for sound value judgments. AI may become extraordinarily skilled at

134. For an argument about a different way in which AI may stunt development of the law, see Daniel Maggen, *Predict and Suspect: The Emergence of Artificial Legal Meaning*, 23 N.C. J.L. & TECH. 67 (2021), available at https://scholarship.law.unc.edu/ncjolt/vol23/iss1/3.

interpreting and applying the law in ways we find credible. That may blind us to the implicit value judgments it makes in assessing how to persuade us. One way to understand this phenomenon is as a variation on the halo effect: our tendency to ascribe to people positive qualities that they do not have because of positive qualities that they do have.[135] We may believe robojudges make sound values judgments because of their skill at opinion writing. A result is that we may miss that the law is listing away from what is right.

Atrophy. Over time we might lose the *ability* to engage in effective legal reasoning. It is a skill. Without practice, skills deteriorate. Without a large stock of human judges—or at least human lawyers—we should not assume that we will remain capable of overseeing or evaluating robojudges.

Complacency. We may also lose the *motivation* to engage in judicial reasoning. It is hard work. It can be stressful. Conscientious judges often struggle to decide cases. If robojudges *seem* to be doing the work of judging well—even if a careful analysis would reveal that they are making serious and accumulating errors—we may not undertake the arduous work necessary to discover the problem and correct course.

Distortion of Our Values. AI may even write judicial opinions that are *designed* to shape our preferences so that they are more predictable. That could taint any feedback loop we develop in an effort to ensure the ongoing quality of AI opinions. In this regard, recall how AI directed social-media users to links that would shape the users' views so that their search habits became more predictable. Stuart Russell suggests that as an explanation for the way in which social media fosters extreme political views.[136] A similar phenomenon could occur in law. AI might end up writing judicial opinions designed to alter our views and preferences so that it can more effectively persuade us. Rather than AI opinions merely reflecting what we find persuasive, they also might *shape* what we find persuasive.

To be sure, the above analysis does not prove that robojudges would be more persuasive opinion writers than human judges or that robojudges would write worse opinions than human judges. There are other possibilities. One is that, all things considered, human judges remain more persuasive than robojudges. Another is that AI, on the whole, will write better opinions—perhaps because its superior technical abilities will more than compensate for its limitations regarding value judgments. Still, we have reason for caution despite AI's potential capacity to persuade—indeed, potentially *because of* its capacity to persuade. AI may turn out to be a devil with a silver tongue.[137]

---

135. The origin of the term is usually traced to E.L. Thorndike, *A Constant Error in Psychological Ratings*, 4 J. APP. PSYCH. 25, 25 (1920).

136. RUSSELL, *supra* note 5, at 8–9.

137. Volokh suggests that we could program robojudges to make persuasive arguments about what is wise or compassionate—and not just about what is legal—if that is what we want. Volokh, *supra* note 3, at 1167. Note, however, that the same points made in the text about law also apply to wisdom and compassion. We should expect AI to be more effective at determining what *seems* wise or compassionate than what *is* wise or

We can encapsulate some of the above analysis by considering two passages in Volokh's article. The first quotes Justice Kagan when she "described the shift from Solicitor General to Supreme Court Justice as shifting 'from persuading nine [Justices] to persuading eight.'"[138] Volokh uses the quotation as evidence that judges write opinions to persuade each other,[139] which seems reasonable.

Note, however, that judges first must decide how they think a case *should* be decided. Justice Kagan presumably would never write an opinion *only* to maximize its odds of winning the other Justices' votes. That would be absurd. If all Justices did that, they would end up in an infinite regress, much like two mirrors facing each other. Each would write an opinion reflecting the anticipated views of the other Justices, which would reflect the anticipated views of the other Justices, ad infinitum. None of them would be making any direct judgments about the law or the facts or how the two relate.

Another way to illustrate this point is by considering a second passage from Volokh, a passage in which he asks a rhetorical question: "What more can we reasonably ask of an opinion drafter—human or AI—than the production of opinions that a blue-ribbon panel of trained observers will accept over the alternatives?"[140] But we can—and do—ask more of some human opinion drafters. We ask judges to try to get their decisions *right*. If AI cannot do that, it may be a poor substitute for us.[141]

CONCLUSION

AI has made tremendous strides at deductive and inductive reasoning. It may in the not-too-distant future improve similarly at abductive reasoning—which could include the kind of common sense that figures prominently in lawyering and judging. If so, that might greatly expand the role of AI in litigation in general and in complex litigation in particular. AI could advise us, advocate for us, help judges assess class action settlements, and propose or impose compromises to resolve legal disputes.

But that does not mean that AI would be able to serve as an effective judge. There is good reason to believe doing so requires a capacity to make reliable judgments about morality or other values. There is also good reason to believe

---

compassionate. A common thread is that wisdom and compassion—like law—are what philosophers sometimes call thick ethical concepts (or perhaps we should say thick normative concepts, to capture other values). *See, e.g.*, PUTNAM, *supra* note 53, at 34–43. The point is: we can no more say with determinacy what is wise or compassionate without making value judgments than we can say what is legal (one might similarly question any sharp distinction between what is wise or compassionate and what is legal.).

138. Volokh, *supra* note 3, at 1149–50, n.49 (quoting Phil Brown, *Associate Justice Elena Kagan Visits NYU Law*, NYU L. COMMENTATOR (Apr. 5, 2016), https://nyulawcommentator.org/2016/04/05/associate-justiceelena-kagan-visits-nyu-law [https://perma.cc/3N32-8KAR] (quoting Justice Kagan)).

139. *Id.*

140. Volokh, *supra* note 3, at 1154.

141. For preliminary thoughts on the implications of these points for legal philosophy, see Joshua P. Davis, *Law Without Mind: AI, Ethics, and Jurisprudence*, 55 CAL. W. L. REV. 165, 212–17 (2018).

that AI will not be capable of making those reliable judgments.[142] That may be true even if we find opinions drafted by robojudges more persuasive than ones drafted by human judges. We should take care about ceding the judiciary in whole or in part to AI. Doing so might corrupt our legal system—rendering the law progressively less just over time.

142. The analysis has assumed that AI will not acquire consciousness. It may. I address reasons to doubt that conscious AI would be capable of reliable judgments about morality or other values in DAVIS, *supra* note 68.