CHIEF JUSTICE ROBOTS

EUGENE VOLOKH†

ABSTRACT

Say an AI program someday passes a Turing test, because it can converse in a way indistinguishable from a human. And say that its developers can then teach it to converse—and even present an extended persuasive argument—in a way indistinguishable from the sort of human we call a "lawyer." The program could thus become an AI brief-writer, capable of regularly winning brief-writing competitions against human lawyers.

Once that happens (if it ever happens), this Essay argues, the same technology can be used to create AI judges, judges that we should accept as no less reliable (and more cost-effective) than human judges. If the software can create persuasive opinions, capable of regularly winning opinion-writing competitions against human judges—and if it can be adequately protected against hacking and similar attacks—we should in principle accept it as a judge, even if the opinions do not stem from human judgment.

† Gary T. Schwartz Professor of Law, UCLA School of Law (volokh@law.ucla.edu). Special thanks to Richard Re for the conversation that helped prompt this essay, and to Jack Balkin, Jane Bambauer, Scott Banister, Ryan Calo, Mark Lemley, Paul Ohm, Ted Parson, David Post, Scott Shapiro, Alicia Solow-Niederman, John Villasenor, and Ze'ev Wurman for their especially detailed and helpful reactions. Thanks also to Howard Bashman, and to audience members at presentations at the University of Haifa, Stanford, UCLA, and Yale.

After coming up with the title for this Article, I learned (unsurprisingly) that others had come up with similar puns before. *See, e.g.*, Sean Braswell, *All Rise for Chief Justice Robot!*, OZY (June 7, 2015), http://www.ozy.com/immodest-proposal/all-rise-for-chief-justice-robot/41131 [https://perma.cc/Q9G2-Q7BP]; Ian Kerr & Carissima Mathen, *Chief Justice John Roberts is a Robot* 8 (2014) (unpublished manuscript), http://robots.law.miami.edu/2014/wp-content/uploads/2013/06/Chief-Justice-John-Roberts-is-a-Robot-March-13-.pdf [https://perma.cc/U5BT-47WJ]. But I couldn't bring myself to give it up.

TABLE OF CONTENTS

INTRODUCTION

How might artificial intelligence change judging? IBM's Watson can beat the top Jeopardy players in answering English-language factual questions.[1] The Watson Debater program can construct persuasive arguments, and indeed squared off against a World Debating Championship grand finalist (Harish Natarajan) in a public debate in February 2019.[2] What would happen if an AI program could write legal briefs and judicial opinions?

To be sure, AI legal analysis is in its infancy; prognoses for it must be highly uncertain. Maybe there will never be an AI program that can write a persuasive legal argument of any complexity.

But it may still be interesting to conduct thought experiments, in the tradition of Alan Turing's famous speculation about artificial intelligence, about what might happen if such a program could be written.[3] Say a program passes a Turing Test, meaning that it can converse in a way indistinguishable from a human. Perhaps it can then converse—or even present an extended persuasive argument—in a way indistinguishable from the sort of human we call a "lawyer," and then perhaps in a way indistinguishable from a judge.

In this Article, I discuss in more detail such thought experiments and introduce four principles—perhaps obvious to many readers, but likely controversial to some—that should guide our thinking on this subject:

1. *Consider the Output, Not the Method.* When we're asking whether something is intelligent enough to do a certain task, the question shouldn't be whether we recognize its reasoning processes as in-

[1] John Markoff, *Computer Wins on 'Jeopardy!': Trivial, It's Not*, N.Y. TIMES (Feb. 16, 2011), https://www.nytimes.com/2011/02/17/science/17jeopardy-watson.html [https://perma.cc/D6SG-X4NF].

[2] *IBM Project Debater*, IQ2US (Feb. 11, 2019), https://www.intelligencesquaredus.org/debates/ibm-project-debater [https://perma.cc/R99H-8SCG]. In my judgment, Project Debater didn't beat the human expert, and the audience seemed to agree; but the AI presented the sort of argument that a good human debater might give--and it's likely to get better still as the software is further refined.

[3] Alan M. Turing, *Computing Machinery and Intelligence*, 59 MIND 433 (1950). People have likewise been thinking about artificial intelligence and the law for many decades now. *See, e.g.*, Bruce G. Buchanan & Thomas E. Headrick, *Some Speculation About Artificial Intelligence and Legal Reasoning*, 23 STAN. L. REV. 40 (1970); Anthony D'Amato, *Can/Should Computers Replace Judges?*, 11 GA. L. REV. 1277 (1977); L. Thorne McCarty, *Reflections on "Taxman": An Experiment in Artificial Intelligence and Legal Reasoning*, 90 HARV. L. REV. 837 (1977).

telligent in some inherent sense. Rather, it should be whether the output of those processes provides what we need.[4]

If an entity performs medical diagnoses reliably enough, it's intelligent enough to be a good diagnostician, whether it is a human being or a computer. We might call it "intelligent," or we might not. But, one way or the other, we should use it. Likewise, if an entity writes judicial opinions well enough—more, shortly, on what "well" means here—it's intelligent enough to be a good AI judge. (Mere handing down of decisions, I expect, would not be enough. To be credible, AI judges, even more than other judges, would have to offer explanatory opinions and not just bottom-line results.)

This, of course, is reminiscent of the observation at the heart of the Turing Test: if a computer can reliably imitate the responses of a human—the quintessential thinking creature, in our experience—in a way that other humans cannot tell it apart from a human, the computer can reasonably be said to "think."[5] Whatever goes on under the hood, thinking is as thinking does.

The same should be true for judging. If a system reliably yields opinions that we view as sound, we should accept it, without insisting on some predetermined structure for how the opinions are produced.[6] Such a change would likely require eventual changes to the federal and state constitutions.[7] But, if I am right, and if the technology passes the tests I describe, then such changes could indeed be made.

2. *Compare the Results to Results Reached by Humans*. The way to practically evaluate results is the Modified John Henry Test, a competition in which a computer program is arrayed against, say, ten average performers in some field—medical diagnosis, translation, or what have you.[8] All the performers would then be asked to execute, say, ten different tasks—for instance, the translation of ten different passages.

[4] Think of this as the Reverse Forrest Gump Principle: "Intelligent is as intelligent does."

[5] Turing, *supra* note 3, at 434.

[6] *See* Harry Surden, *Machine Learning and Law*, 89 WASH. L. REV. 87, 95 (2014) (describing such an "outcome-oriented view of intelligence"). This is as true for decisions about procedural rules as about substantive rules: Even if the legal question before the judge is whether certain procedures should be followed, we should evaluate the judge's opinions, and not whether the judge arrives at the opinions through traditional human reasoning or through a computer program.

[7] *See infra* note 72 and accompanying text.

[8] The name, of course, is borrowed from the famous folk story of John Henry, a "steel-driving man" whose job was to hammer out holes for explosives used in railroad tunnel blasting. When the steam drill, which would do the task automatically, was offered as a replacement for manual labor, the legend goes that John Henry offered to match himself against the drill.

Sometimes this performance can be measured objectively. Often, it can't be, so we would need a panel of, say, ten human judges who are known to be experts in the subject—for example, experienced doctors or fluent speakers of the two languages involved in a translation. Those judges should evaluate everyone's performance without knowing which participant is a computer and which is human.

If the computer performs at least as well as the average performer, then the computer passes the Modified John Henry Test.[9] We can call it "intelligent" enough in its field. Or, more to the point, we can say that it is an adequate substitute for humans.[10]

I label the test the *Modified* John Henry Test because of what I call the Ordinary Schlub Criterion. As I noted above, a computer doesn't have to match the best of the best; it just has to match the performance of the average person whom we are considering replacing.[11]

Self-driving cars, to offer an analogy, do not have to be perfect to be useful—they just have to match the quality of ordinary drivers, and we ordinary drivers don't set that high a bar. Likewise, translation software just has to match the quality of the typical translator who would be hired in its stead.[12] Indeed, over time we can expect self-driving cars

John Henry won—"made his fifteen feet / The steam drill only made nine." HARRY BELAFONTE, JOHN HENRY (1954). But in the long run, the steam drill won, even if it had to wait until Steam Drill 2.0, and steel-driving men are history. (John Henry also "drove so hard he broke his poor heart / And he laid down his hammer and he died," *id.*; fortunately, that need not happen in our test.)

[9] This doesn't require a unanimous judgment on the part of the panel; depending on how cautious we want to be, we might be satisfied with a majority judgment, a supermajority judgment, or some other decision rule.

[10] In some contexts, of course, automation may be better even if it's not as effective—for instance, it may be cheaper and thus more *cost*-effective. But if it's cheaper *and* at least as effective, then it would be pretty clearly superior. *See* Ian Kerr & Carissima Mathen, *Chief Justice John Roberts is a Robot* 8 (2014) (unpublished manuscript), http://robots.law.miami.edu/2014/wp-content/uploads/2013/06/Chief-Justice-John-Roberts-is-a-Robot-March-13-.pdf [https://perma.cc/U5BT-47WJ] (describing, though ultimately not endorsing, the view that "[i]f AI does become better and more reliable than human professionals at carrying out [certain] tasks . . . evidence-based reasoning will demand that we choose AI over human experts based on its better record of success").

[11] *See generally* Jane R. Bambauer, *Dr. Robot*, 51 UC DAVIS L. REV. 383 (2017) (making the same point with regard to the proper baseline for evaluating AI physicians).

[12] Carl Sagan observed that no computer program "is adequate for psychiatric use today [in 1975], but the same can be remarked about some human psychotherapists." D'Amato, *supra* note 3, at 1284 (quoting Carl Sagan, *Comment*, 84 NAT. HIST. 10 (1975)). The question is never whether a proposed computer solution is imperfect; it's whether it's good enough compared to the alternative.

and translation software to keep improving as the technology advances; the humans' average, on the other hand, is not likely to improve, or at least to improve as fast. But even without such constant improvement, once machine workers are as good as the average human workers, they will generally be good enough for the job.

Indeed, in the John Henry story, Henry's challenge was practically pointless, though emotionally fulfilling. Even if John Henry hadn't laid down his hammer and died at the end, he would have just shown that a team of John Henrys would beat a team of steam drills. But precisely because John Henry was so unusually mighty, the railroad couldn't hire a team of workers like him. The railroad only needed something that was faster than the average team—or, more precisely, more cost effective than the average team.[13] Likewise for other technologies: to be superior, they merely need to beat the human average.

Now, in some contexts, the ordinary schlub may be not so schlubby. If you work for a large company with billions at stake in some deal, you might hire first-rate translators—expensive, but you can afford them. Before you replace those translators with computer programs, you would want to make sure that the program beats the average translator of the class that you hire. Likewise, prospective AI Supreme Court Justices should be measured against the quality of the average candidates for the job—generally experienced, respected appellate judges—rather than against the quality of the average candidate for state trial court.[14]

Nonetheless, the principle is the same: the program needs to be better than the average of the relevant pool. It doesn't need to be perfect, because the humans it would replace aren't perfect. And because such a program is also likely to be much cheaper, quicker, and less subject to certain forms of bias, it promises to make the legal system not only more efficient but also fairer and more accessible to poor and middle-class litigants.

3. *Use Persuasion as the Criterion for Comparison—for AI Judges as Well as for AI Brief-Writers*. Of course, if there is a competition, we

[13] It didn't matter if he won, if he lived, or if he'd run.
They changed the way his job was done. Labor costs were high.
That new machine was cheap as hell and only John would work as well,
So they left him laying where he fell the day John Henry died.

DRIVE-BY TRUCKERS, THE DAY JOHN HENRY DIED (2004).

[14] *See* Bambauer, *supra* note 11, at 390.

need to establish the criteria on which the competitors will be measured. Would we look at which judges' decisions are most rational? Wisest? Most compassionate?

I want to suggest a simple but encompassing criterion, at least for AI judges' judgment about law and about the application of law to fact: persuasion. This criterion is particularly apt when evaluating AI brief-writer lawyers. After all, when we hire a lawyer to write a brief, we want the lawyer to persuade—the lawyer's reasonableness, perceived wisdom, and appeals to compassion are effective only insofar as they persuade. But persuasion is also an apt criterion, I will argue, for those lawyers whom we call judges. (The test for evaluation of facts, though, whether by AI judges, AI judicial staff attorneys, or AI jurors, would be different; I discuss that in Part IV.)

If we can create an AI brief-writer that can persuade, we can create an AI judge that can (1) construct persuasive arguments that support the various possible results in the case, and then (2) choose from all those arguments the one that is most persuasive, and thus the result that can be most persuasively supported. And if the Henry Test evaluator panelists are persuaded by the argument for that result, that means they have concluded the result is correct. This connection between AI brief-writing and AI judging is likely the most controversial claim in the paper.

4. *Promote AIs from First-Draft-Writers to Decisionmakers.* My argument starts with projects that are less controversial than AI judges. I begin by talking about what should be a broadly accepted and early form of AI automation of the legal process: the use of AI interpreters to translate for non-English-speaking witnesses and parties.[15] I then turn to AI brief-writing lawyers—software that is much harder to create, of course, but one that should likewise be broadly accepted, if it works.[16]

From there, I argue that AI judicial staff attorneys that draft proposed opinions for judges to review—as well as AI magistrate judges that write reports and recommendations rather than making final decisions—would be as legitimate and useful as other AI lawyers (again, assuming they work).[17] I also discuss AIs that could help in judicial fact-finding, rather than just law application.[18]

[15] *See infra* Part I.

[16] *See infra* Part II.

[17] *See infra* Part III.

[18] *See infra* Part IV.

And these AI judicial staff attorneys and magistrates offer the foundation for the next step, which I call the AI Promotion: If we find that, for instance, AI staff attorneys consistently write draft opinions that persuade judges to adopt them, then it would make sense to let the AI make the decision itself—indeed, that can avoid some of the problems stemming from the human prejudices of human judges.[19] I also discuss the possible AI prejudices of AI judges, and how they can be combated.[20]

Just as we may promote associates to partners, or some magistrate judges to district judges, when we conclude that their judgment is trustworthy enough, so we may promote AIs from assistants to decisionmakers. I also elaborate on the AI Promotion as to jurors,[21] and finally move on to the title of this Article: AI judges as law developers.[22]

Indeed, the heart of my assertion in this Article is this: the problem of creating an AI judge that we can use for legal decisions[23] is not materially more complicated than the problem of creating an AI brief-writer that we can use to make legal arguments.[24] The AI brief-writer may practically be extremely hard to create. But if it is created, there should be little conceptual reason to balk at applying the same technology to AI judges within the guidelines set forth below. Instead, our focus should be on practical concerns, especially about possible hacking of the AI judge programs, and possible exploitation of unexpected glitches in those programs; I discuss that in Part V.C.3.

This, of course, is likely to be a counterintuitive argument, so I try to take it in steps, starting with the least controversial uses of AI: courtroom interpreters (Part I), brief-writing lawyers (Part II), law clerks (Part III), and fact-finding assistants that advise judges on evaluating the facts, much as law clerks do as to the law (Part IV). Then I shift from assistants to actual AI judges (Part V), possible AI jurors (Part VI), and finally AI judges that develop the law rather than just apply it (Part VII); those three parts are where I argue that it makes sense to actually give AIs decision-making authority. It would be a startling

[19] *See infra* Part V.

[20] *See infra* Part V.C.

[21] *See infra* Part VI.

[22] *See infra* Part VII.

[23] Fact-finding is a somewhat different matter. *See infra* Part IV.

[24] The slight extra complexity is discussed in Part V.A: an AI judge also needs to have a module that compares two possible opinions and determines which of them is more likely to be persuasive.

step, but—assuming that the technology is adequate and that we can avoid an intolerable level of security vulnerabilities—a sound one.

I. COURTROOM INTERPRETERS

Let's begin with a mundane but critical aspect of the legal system: interpreting for parties and witnesses who don't speak English. Interpreters are expensive, and good interpreters are often in short supply in the court system. And because to err is human, interpreters sometimes make mistakes.[25]

Courtroom interpreting seems likely to be automatable. Current translation software isn't good enough yet, but it has been improving quickly, and really good translation software looks like it's coming fairly soon.[26] Add voice recognition and voice synthesis.[27] Design the software so it can ask clarifying questions, or accept corrections, like a good human translator can.[28] And then many users of interpreters—including courts—will be interested in switching.[29]

Of course, they shouldn't switch until the software is ready. But once the software passes the Henry Test, what reason would there be

[25] *See, e.g.*, Anna M. Nápoles, Jasmine Santoyo-Olsson, Leah S. Karliner, Steven E. Gregorich & Eliseo J. Pérez-Stable, *Inaccurate Language Interpretation and Its Clinical Significance in the Medical Encounters of Spanish-Speaking Latinos*, 53 MED. CARE 940, 947 (2015) (reporting that even professional interpreters had an error rate of 25%). This problem is especially severe when, as often happens, cost constraints keep courts from hiring enough highly competent interpreters. *See, e.g.*, United States v. Mosquera, 816 F. Supp. 168, 171 (E.D.N.Y. 1993); Beth Gottesman Lindie, *Inadequate Interpreting Services in Courts and the Rules of Admissibility of Testimony on Extrajudicial Interpretations*, 48 U. MIAMI L. REV. 399, 410 (1993) ("The inability to find and employ qualified, if not certified, interpreters has led to the widespread use of unqualified and incompetent individuals as interpreters."). For some examples, see *Perez-Lastor v. INS*, 208 F.3d 773 (9th Cir. 2000); Geesche Jacobsen, *Trial Aborted After Juror Criticises Interpreter*, SYDNEY MORNING-HERALD (Nov. 8, 2011, 3:00 AM), https://www.smh.com.au/national/nsw/trial-aborted-after-juror-criticises-interpreter-20111107-1n3tn.html [https://perma.cc/GFZ3-8KNN].

[26] Karen Turner, *Google Translate Is Getting Really, Really Accurate*, WASH. POST (Oct. 3, 2016), https://www.washingtonpost.com/news/innovations/wp/2016/10/03/google-translate-is-getting-really-really-accurate/?noredirect=on&utm_term=.e9b4e2b0205c [https://perma.cc/C9KV-DCUX].

[27] Jon Russell, *Google Translate Now Does Real-Time Voice and Sign Translations on Mobile*, TECHCRUNCH (Jan. 14, 2015), https://techcrunch.com/2015/01/14/amaaaaaazing [https://perma.cc/E62W-U4FP]. This software is still an early version, but it will keep getting better.

[28] *See, e.g.*, METTE RUDVIN & ELENA TOMASSINI, INTERPRETING IN THE COMMUNITY AND WORKPLACE: A PRACTICAL TEACHING GUIDE 62–63 (2011).

[29] Note that this is so whether or not one views translation as "true artificial intelligence"—except insofar as the label may be seen as a promotional tool, or for that matter as a promotional handicap. The question is whether the translation software is effective, whether or not it is "intelligent."

not to use it? True, the automated interpreter will make mistakes. But so do human interpreters; if the interpreter passes the Henry Test, then its mistakes are no more frequent than the human mistakes.[30]

One might worry that the software may be biased in some ways, whether because of its human designers' plans or because of some "emergent property" that the designers didn't foresee. What if, for instance, the software is less reliable for people who speak with certain accents? Or what if it chooses the more incriminating alternatives when a word is ambiguous?

But, again, we humans likewise have biases, deliberate or subconscious. Ultimately, the proof of the pudding is in the eating. If, in the minds of whatever experts we place on the Henry Test evaluation panel—experts who are not told whether an interpreter is human or automated—the AI interpreter is at least as reliable and at least as bias-free as the average human interpreter, then the AI interpreter should be adopted.

We may still much regret the AI interpreter's errors and biases, and we should try to minimize them by further improving the technology.[31] But we shouldn't let the perfect be the enemy of the good. We should accept an AI interpreter as soon as it becomes at least as competent as the human interpreter, even if it occasionally errs.

Surely that's the way that businesses who hire interpreters would operate. And there's no good reason for courts to take a different view.

II. LAWYERS AS BRIEF-WRITERS

Now let's imagine someone designs something that's much further away than good interpretation software: a program that writes briefs.

[30] Allison Linn, *Historic Achievement: Microsoft Researchers Reach Human Parity in Conversational Speech Recognition*, MICROSOFT AI BLOG (Oct. 18, 2016), https://blogs.microsoft.com/ai/historic-achievement-microsoft-researchers-reach-human-parity-conversational-speech-recognition [https://perma.cc/6DXN-YGYU]. To be precise, we want the mistakes to be no more frequent and no more serious; a 10% rate of mistakes is worse if the mistakes radically misrepresent what is being said (e.g., translate "I didn't do it" as "I did it") than if the mistakes are comparatively minor. In principle, we might want to optimize for the percentage of mistakes weighted by their comparative seriousness.

[31] It may help that the prospects for improving AI interpreter performance—better algorithms and the like—will likely be much better than those for improving the performance of the average human interpreter.

It takes all the record documents, figures out the legal issues and arguments that would be considered relevant by a judge in the particular jurisdiction, and produces a brief, whether trial level or appellate.[32]

It is possible that AI will never get good enough at this. Processing all the documents—contracts, statutes, precedents, witness testimony, emails—in the way needed to construct a persuasive legal argument might be too hard a task.[33]

That is especially so since persuasive legal argument must be not only about applying clear rules—was a document signed? did it need to be?—but also about vaguer standards, such as "reasonableness" or whether the "probative value [of evidence] is substantially outweighed by a danger of . . . unfair prejudice."[34] And, perhaps hardest of all, such an argument has to deal with questions of credibility and factual inference: Which witness is telling the truth? Is someone's story internally consistent? Are certain allegations so improbable that we should require an especially great deal of evidence for them?

But suppose that years from now, some company says that it has succeeded in solving these problems. After all, if AIs ever pass the Turing Test, that means they will be able to converse like ordinary humans do, at least in writing. Imagine, then, AIs that can converse like lawyers do, and when asked to explain why their client should prevail on some issue can offer an answer—indeed, an extended, brief-long answer—on the subject.

Now, say you work for a business that is tired of paying lawyers top dollar for this work, so you're intrigued. But you obviously want quality legal work, and you wonder whether you're going to get it from the AI.

Again, you would need to conduct a Henry Test, with the criterion being *persuasion*. Hire ten lawyers to write ten briefs each. Have the software write briefs on the same issues. Hire a panel of ten retired judges whom you trust to tell you which persuades them most, without their knowing who wrote what. If the AI is at least as good as the average human brief-writer, why would you go with the more expensive

[32] By analogy to Chief Justice Robots, Neal KatyAI? "[P]erhaps artificially intelligent appellate advocates will play the role of the steam hammer in a folk tale about how Neal Katyal or Paul Clement was a brief-writing man who died slumped over the podium having defeated his computerized opponent." Travis Ramey, *Appellate A.I.*, APP. ISSUES, Nov. 2017, at 14, 19.

[33] I am not suggesting that current "machine learning" tools—for instance, ones that help with document review, *see, e.g.*, Julie Sobowale, *How Artificial Intelligence Is Transforming the Legal Profession*, ABA J., Apr. 2016, at 1, 1—are anywhere near what is required for this.

[34] FED. R. EVID. 403.

and no-more-effective human, when you can use the cheaper and possibly better (or at least equal) computer program?

The test will be expensive, but it can be run once on behalf of many potential clients. The software developer will likely pay to have the test run by a credible, impartial organization—perhaps a panel of retired judges who are paid a flat fee up front. And if the software is at least as good as the average of the human lawyers, clients can save a vast amount of money going forward.[35]

To be sure, some could ask what is "really" going on in the process. Is the software "understanding" the precedents, the record documents, the policy arguments? Is it truly engaging in "analogical reasoning"?[36] Is it exercising "legal judgment"? Is it "intelligent"?

Yet intelligent is as intelligent does. There is nothing mystical about the result you, as a prospective client, seek—you want the AI brief-writer to persuade your target audience to make the decision you want. That could be a monumentally difficult design problem. But if the software can accomplish that, that's all you need.

It might be better if we try to avoid, when possible, the language that we too closely associate with human minds. Instead of the software "understanding" some documents, we might be better off talking about the software determining how the legal rules can be persuasively argued to apply to the contents of the document. Instead of "intelligent," we can just say "effective."[37]

Of course, what persuades turns on the identity of the person you are trying to persuade. The software would have to be programmed

[35] Indeed, some clients might be satisfied with software that is not even as good as the average of the human lawyers if it is sufficiently cheaper. But for the sake of simplicity, I focus only on the criterion being persuasiveness rather than (more precisely) persuasiveness relative to cost.

[36] Cass Sunstein, for instance, argues that analogical reasoning may be especially hard to program, because it requires arguments about value judgments. Cass Sunstein, *Of Artificial Intelligence and Legal Reasoning*, 8 U. CHI. L. SCH. ROUNDTABLE 29, 33–34 (2001). That might indeed hinder the development of good AI brief-writers because they would have to make arguments for making such judgments, and these arguments would require the software to go beyond the four corners of the precedent and the current case's fact pattern.

[37] As Edsger Dijkstra famously put it—as it happens, referring to Turing's work—"[t]he question of whether Machines Can Think . . . is about as relevant as the question of whether Submarines Can Swim." THE YALE BOOK OF QUOTATIONS 205 (Fred R. Shapiro ed., 2006); Edsger W. Dijkstra, Speech to Association for Computing Machinery 1984 South Central Regional Conference: The Threats to Computing Science (Nov. 16–18, 1984), https://www.cs.utexas.edu/users/EWD/transcriptions/EWD08xx/EWD898.html [https://perma.cc/CDS6-MRDA]. Submarines can propel themselves through the water, whether or not one calls it "swimming." If an AI can produce persuasive arguments, that is what matters, not whether this is done through a process we normally call "thinking," "reasoning," or "intelligence."

accordingly, just as we teach young lawyers to act accordingly. The program would obviously have to recognize that different jurisdictions have different substantive rules. It might also recognize that different localities, and even different people, have different rhetorical preferences.

Yet the question should solely be whether we can develop software that is capable of this.[38] If at some point, we can do this (a big "if," I realize), then we would be foolish to forgo the cost savings—and eventually the greater persuasion—that the software can offer. Indeed, for many people, even a not very good AI lawyer may be better than no lawyer at all, especially if that is all they can afford.[39] Advancing technology has helped put many formerly expensive goods—clothing, food, entertainment, and more—within reach of the poor.[40] Realistically, the only way we are likely to sharply increase access to expensive services, such as lawyering, is through technology.

If clients aren't comfortable with just relying on the AI software, they can use what we might call the AI Associate model: they can have the AI software write a first draft of the brief and then have an experienced human lawyer—the equivalent of the modern partner or in-house counsel—review and edit it, for a fraction of the cost of writing it from scratch. This is similar to how many translators are already using machine translation,[41] though this process wouldn't work as well for real-time interpretation.

But even this review process might at some point become obsolete. Here, we can keep running the Henry Test: we can compare the unedited AI brief-writer with the combination of the AI brief-writer and a human editor to see whether there is a material difference in persuasion, measured, again, by a panel of judges who don't know which submission is which. If at some point, there is no measurable difference, then even the cost of human editing—though much less than the cost of human beginning-to-end writing—might no longer be justifiable.[42]

[38] Or, to be pedantic, asking the questions that match Jeopardy answers.

[39] *See* D'Amato, *supra* note 3, at 1286.

[40] *See, e.g.*, *Make It Cheaper, and Cheaper*, ECONOMIST (Dec. 11, 2003), https://www.economist.com/special-report/2003/12/11/make-it-cheaper-and-cheaper [https://perma.cc/88AA-9333] (discussing how technology has made food cheaper).

[41] *See, e.g.*, Rebecca Fiederer & Sharon O'Brien, *Quality and Machine Translation: A Realistic Objective?*, J. SPECIALISED TRANSLATION, no. 11, Jan. 2009, at 52, 52.

[42] Indeed, if the AI associates entirely replace human associates, then some decades later all the human lawyers will have retired or died, and there may be no one to provide legal editing

Naturally, there will be political resistance to this by human lawyers, who may rightly worry that the AIs will take away human jobs. And human lawyers have considerable power, through their control of state bars, to suppress competition.[43]

But big businesses that are tired of paying vast sums for attorney fees have considerable power, too. I doubt that even lawyers will long be able to resist calls to allow such businesses to use the latest labor-saving technology. And once the Microsofts and GMs of the world can use AI brief-writers, less-powerful small businesses and consumers will likely be able to use the same technology. Indeed, I expect the prospect of AI brief-writers to be the main impetus for investing in developing AI legal-writing technology, precisely because there is such a potential for savings for businesses here—and thus such a potential for profit for AI developers.

III. JUDICIAL STAFF ATTORNEYS / ADVISORY MAGISTRATES

A. Judicial Staff Attorney 1.0: Writing the Opinion as Instructed by the Judge

Now, let's turn to a particular kind of lawyer: the staff attorneys who work for judges, sometimes called career law clerks.[44] One job of a staff attorney is to draft opinions that the judge can then edit and sign, usually based on a big picture outline the judge gives: "I think the statute is unconstitutional because it is overinclusive with respect to the government interest, see case X and case Y. Write an opinion that elaborates on this."

Suppose someone takes the hypothetical AI Brief-Writer and turns it into an AI Staff Attorney that is supposed to write draft opinions that persuade, rather than briefs that persuade.[45] This modification should not be a difficult task. True, the formatting and the tone of the output needs to be changed, and counterarguments should probably be

of the AI associates' work (as opposed to general rhetorical editing that skilled human nonlawyer persuaders can provide).

[43] *See, e.g.*, Deborah L. Rhode & Lucy Buford Ricca, *Protecting the Profession or the Public? Rethinking Unauthorized-Practice Enforcement*, 82 FORDHAM L. REV. 2587, 2605 (2014).

[44] The more familiar judicial law clerks, usually recent graduates appointed for a year or two, are a form of staff attorney, but I set them aside as analogies because their lack of experience may lead judges to especially closely supervise them.

[45] *See* Kerr & Mathen, *supra* note 10, at 7. As noted above, even existing, imperfect machine translation systems are often used to provide first drafts that human translators can then correct.

treated more prominently and less dismissively. But, ultimately, the general nature of the task should be similar.

The judge can then instruct the AI Brief-Writer to write an opinion that comes out a particular way; the judge would then review the opinion and edit it as necessary. Or if the judge is unsatisfied, the judge can ask for an opinion that reaches the opposite result (or at least a different result); perhaps that opinion would be better.

AI theorists have noted that at least early AIs will be aimed at helping human decisionmakers—say, human doctors who are using AI diagnostic tools—rather than supplanting them.[46] The AI Staff Attorney can likewise help judges make their decisions, just as human staff attorneys do today. If the program passes the Henry Test because of its ability to write opinions that persuade the evaluation panel, then the court system would benefit from using the program instead of the more expensive, slower human staff attorneys.

Some might balk at characterizing a judicial opinion as an attempt to persuade, but I think that is indeed a sound characterization, and indeed what a judge should want from a staff attorney or law clerk.

First, judicial opinions may try to persuade the parties—and the public, to the extent the public is interested in the case—that their answers are right, or at least that the parties' arguments have been thoughtfully considered.[47] Among other things, this sort of persuasion is important for maintaining the legitimacy of the legal system; the persuasive arguments may be most effective when they are framed as merely expressing the law, rather than as an attempt to persuade, but that framing is itself a means of persuasion.

Second, judges on multimember courts often write opinions aimed at persuading other judges to join the opinion.[48] Justice Kagan, for instance, has described the shift from Solicitor General to Supreme

[46] *See, e.g.*, Surden, *supra* note 6, at 101; *see also* Frank Pasquale, *A Rule of Persons, Not Machines: The Limits of Legal Automation*, 87 GEO. WASH. L. REV. 1, 46–49 (2019) (discussing such "augmentation" of human intelligence by computer assistants, though arguing against use of AI to supplant human decisionmakers).

[47] *See, e.g.*, Brown v. Gamm, 525 F.2d 60, 61 n.2 (8th Cir. 1975) ("The district court's opinion might so persuade the parties that appeal could be avoided entirely . . . ."); Wilson Huhn, *The Stages of Legal Reasoning: Formalism, Analogy, and Realism*, 48 VILL. L. REV. 305, 331 (2003) ("[T]he authors of these [judicial] opinions bear a professional obligation to persuade the parties, the profession and society that the decisions are dictated by law."); James G. Wilson, *The Role of Public Opinion in Constitutional Interpretation*, 1993 BYU L. REV. 1037, 1068 ("[Early American courts] wrote their opinions to persuade both the litigating parties and the public of the correctness of their decisions.").

[48] Brett G. Scharffs, *The Character of Legal Reasoning*, 61 WASH. & LEE L. REV. 733, 746–

Court Justice as shifting "from persuading nine [Justices] to persuading eight."[49]

Third, especially if judges' opinions are subject to de novo appellate review, the judges may try to persuade an appellate court to affirm. As it happens, trial courts in Pennsylvania actually write their opinions in precisely this voice, for instance:

> Appellant, John Brown, appeals from this court's Order of February 23, 2017, granting judgment for possession of the property entered in favor of Appellee. . . .
>
> FACTUAL AND PROCEDURAL HISTORY. . . .
>
> DISCUSSION. . . .
>
> CONCLUSION
>
> For all of the reasons stated above, this court's order should be affirmed.[50]

The judges aren't trying to persuade the appellate court on behalf of a client, but they are still trying to persuade.[51]

47 (2004).

[49] *See* Phil Brown, *Associate Justice Elena Kagan Visits NYU Law*, NYU L. COMMENTATOR (Apr. 5, 2016), https://nyulawcommentator.org/2016/04/05/associate-justice-elena-kagan-visits-nyu-law [https://perma.cc/3N32-8KAR] (quoting Justice Kagan).

[50] Help PA IV, LP. v. Brown, No. 161102512, 2017 WL 3077942, at *1–2 (Pa. Ct. C.P. Phila. Cty. July 12, 2017). Search for "should be affirmed" in Pennsylvania courts in Westlaw, limit the result to Trial Court Orders, and you'll see a lot of this. Two other examples: "As discussed herein below, the Order should be affirmed." Jarrett v. Consol. Rail Corp., No. 15021295, 2017 WL 3077936, at *1 (Pa. Ct. C.P. Phila. Cty. June 29, 2017). "The undersigned has impartially and dispassionately reviewed the entirety of these proceedings from the inception and respectfully suggests that the trial was fair and the verdict was just, hence the judgment of sentence should be affirmed." Commonwealth v. Kane, No. CP-46-CR-6239-2015, 2017 WL 2366702, at *103 (Pa. Ct. C.P. Montgomery Cty. Mar. 2, 2017).

I suspect this framing stems partly from PA. R. APP. P. 1925, which is titled "Opinion in Support of Order," and which requires judges to write opinions only once a notice of appeal is filed, PA. R. APP. P. 1925(a)(1)—at that point, it is especially tempting to view the opinion as addressing the appellate court. On the other hand, the first opinion I could find that fits this pattern was from 1983, Appeal of Senft from the Decision of the Lower Merion Twp. Zoning Hearing Bd., 31 Pa. D. & C.3d 578 (Pa. Ct. C.P. 1983), which happened eight years after Rule 1925 was adopted. Perhaps there is a different reason for the practice, or perhaps the Rule influenced the practice but only after some years of experience under it.

[51] I set aside here judges who are trying to persuade higher-court judges—or future judges on the same court or other courts—to adopt a new legal principle. Part VII discusses judges engaged in developing legal rules; here, the focus is on judges finding facts or applying legal principles.

Of course, governments might be reluctant to invest the massive amounts of money needed to develop AI staff attorneys (or, eventually, AI judges) from scratch. But once the AI brief-writers are developed, and paid for by the business clients described in the previous section, adapting them to be AI staff attorneys shouldn't be expensive. Legislatures that want to cut the costs of the court system would then be able to easily justify helping pay these adaptation costs—whether directly or through buying the software that commercial developers have invested in developing—given the resulting savings of the salaries of human staff attorneys.[52]

B. Judicial Staff Attorney 2.0: Proposing the Result to the Judge

Of course, staff attorneys sometimes do more than just write an opinion to reach the result the judge has reached. Sometimes, they recommend a particular result to the judge.

This is common when staff attorneys write bench memos that are given to judges before the judges even read the briefs and hear oral argument. Likewise, some cases that are seen as simple enough are sometimes routed to staff attorneys who present a special screening panel of judges with a draft decision, whether orally or in writing.[53] Federal magistrate judges play a similar role when they write a report and recommendation that is to be reviewed by a district judge.[54] If they are to do the same, AI Staff Attorneys must go beyond the AI Brief-Writer technology.

But not too far beyond. Here's a first cut at the problem: the judge can ask the AI Staff Attorney to prepare a separate opinion for each possible outcome. The judge could then read the opinions, see which one persuades him, and then adopt that one.

Now a second cut: if AIs can be programmed to write persuasive briefs, presumably they can be programmed to evaluate—with considerable accuracy, even if not perfect accuracy—which of the briefs is most likely to persuade. If that is so, then AI Staff Attorney 2.0 can automatically compose briefs supporting all the plausible outcomes in a case and then simply choose the one that, according to its algorithm,

[52] Thanks to David Edelsohn for stressing the importance of this issue.

[53] In the Ninth Circuit, for instance, these are labeled "one-weight" cases. *See* Judge Morgan Christen, *Introduction*, 43 GOLDEN GATE U. L. REV. 1, 4 (2013) (describing one-weight cases as typically having one or two issues on which the court has directly controlling authority).

[54] I am speaking here of magistrate judges acting solely in their advisory capacity, not of the situations where the magistrate judges actually make binding decisions.

scores highest on the likely persuasion meter. That is the draft opinion that the judge will review.[55]

If the likely persuasion meter is sound, then, by hypothesis, the judge will be persuaded to adopt the opinion. Indeed, here, too, we would wait to accept AI Staff Attorney 2.0 until it passes the Henry Test by yielding results that persuade a panel of evaluators at least as often as the results produced by human staff attorneys. Once a program does pass such a test, then it would be a reliable, cost-effective, and quick alternative to human staff attorneys.

C. Why Persuasion Rather than Correctness?

I keep talking about an opinion that persuades rather than an opinion that is fair, wise, or correct. Is that sound?

The problem, of course, is that there are multiple judgments about what is fair, wise, or correct. And these judgments don't just vary among observers—in many cases, the same observer can say that there are several positions that are defensible as correct. Indeed, Legal Realists would likely doubt any legal project aimed at always finding a legal "right answer,"[56] especially when we get to standards such as reasonableness.

But for any legal opinion presented to an evaluator, the evaluator has to decide: Did the opinion persuade me? Did it lead me to conclude that the result is legally correct, however I might understand "correctness" for this particular legal question?[57] Did it lead me to agree with any discretionary judgment calls it had to make, such as whether some evidence would be substantially more unfairly prejudicial than relevant, or what sentence is proper given the circumstances?[58]

[55] For those who like formulas, let R be the set of possible results that are available in a particular case, let $M(r)$ be the most persuasive opinion for any particular result r, and let $P(o)$ be the persuasiveness of that opinion o (as always, persuasiveness to a particular evaluative body). The result that AI Staff Attorney 2.0 would choose would be the result that could be most persuasively defended:

$$a \in R \mid (\nexists b \in R \mid P(M(b)) > P(M(a)))$$

And the opinion that it would write for that result would be $M(a)$.

[56] In John Harrison's humorous locution, some legal scholars may say they subscribe to "naïve right-answerism," but that is to distinguish themselves from a mainstream that does not. *See* Symposium, *Discussion: The Role of the Legislative and Executive Branches in Interpreting the Constitution*, 73 CORNELL L. REV. 386, 386 (1988) (discussing the participants' support for right-answerism).

[57] *See infra* note 153 and accompanying text.

[58] I do not mean persuading in the sense of changing the evaluator's underlying beliefs, but simply persuading in the sense of leading the evaluator to agree with the proposed opinion. Under

All these factors thus get subsumed within persuasion, and without any need to decide what *the* supposedly correct answer is. In ordinary litigation, the winning side is the side that persuades the judge, even if there is no logical way to prove that the winning answer is correct. Likewise, when we're running a Henry Test to evaluate judges, we should ask simply which candidate most often persuades the evaluators.

D. Persuasion Anxiety

The word "persuasion" doesn't always have a positive connotation. Some persuasion is seen as manipulative—as a means of getting us to do something against our better judgment and against our true interests.

Cicero, among many others, warned that "eloquence without wisdom is, in most instances, extremely harmful and never beneficial."[59] Likewise, in Jane Austen's novel *Persuasion*,[60] Austen's point seems to be that the heroine's family was wrong to persuade her to reject her true love, and that the heroine was wrong to be persuaded. There, the persuasion risked causing emotional harm by excessive appeal to economic interests. But often we worry about a persuasive speaker—say, a politician or an advertiser—duping people into doing economically irrational things through appeal to emotion.[61]

But recall that the AI Staff Attorney software, like the AI Brief-Writer software, would have to pass a Henry Test administered by retired human judges. These evaluators would possess decades of experience with resisting undue manipulation at the hands of trained persuaders. These retired judges will likely be pretty good at identifying

this definition, if I want you to buy a bowl of tsukemen from me for five dollars, I give you an extended argument for why you should do that, and you agree, then I have persuaded you, even if the persuasion was very easy (I had you at "tsukemen five dollars," since you are hungry, love tsukemen, know that my tsukemen is delicious, and know that five dollars is a good deal), and not just if it took a lot of work (you've just eaten, you had never heard of tsukemen, you've had a bad experience with tsukemen, you were suspicious of my tsukemen, or you at first thought five dollars was a bit much).

[59] Marcus Tullius Cicero, *De Inventione 1.1*, *in* MARCUS TULLIUS CICERO, HOW TO WIN AN ARGUMENT: AN ANCIENT GUIDE TO THE ART OF PERSUASION 10 (James M. May ed. & trans., 2016).

[60] JANE AUSTEN, PERSUASION (Gillian Beer ed., 1998).

[61] Consider cartoonist Scott Adams's characterization—part critical and part admiring—of candidate Donald Trump as a "master persuader." Scott Adams, *The Trump Master Persuader Index and Reading List*, DILBERT (Feb. 18, 2016), http://blog.dilbert.com/2016/02/18/the-trump-master-persuader-index-and-reading-list [https://perma.cc/BWL2-9TQC].

arguments that make legal sense, and at discarding improper appeals to emotion.[62]

Beyond that, if the panel of retired judges is actually persuaded by an argument, then that means they have come to accept its reasoning. 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?

IV. FACT-FINDING ASSISTANTS[63]

We have been talking so far about staff attorneys that draft opinions applying law to facts. But judges also often have to *find* facts—in bench trials, in injunction hearings, in preliminary decisions about the admissibility of evidence, and so on.[64]

Fact-finding is a different matter than law application, and it requires a different set of skills. It may require an ability to accurately evaluate a witness's demeanor as a guide to whether the witness is lying, evasive, or uncertain. It also requires an ability to consider consistencies and inconsistencies in each witness's story, as well as which witnesses and documents are consistent or inconsistent with each other. It requires an ability to evaluate human biases, human perception, and human memory. It is thus possible that good AI fact-finding software may be much harder to write than the AI Brief-Writer or AI Staff Attorney.

On the other hand, AIs may have some advantages over humans here, partly because humans aren't very good at these things. First, there is some reason to think that a person's demeanor does offer some clues about whether the person is telling the truth, but that these clues are too subtle for most humans to pick up. Computers' greater processing speed and attention to detail may enable them to more effectively detect lies.[65]

[62] They may be open to appeals to emotion that are seen as acceptable within the norms of the existing legal culture, but, by hypothesis, those would be the sorts of appeals that are seen as legitimate persuasion, rather than illegitimate.

[63] I am particularly indebted in this Section to Jane Bambauer.

[64] *See, e.g.*, FED. R. EVID. 104(a) ("[A] court must decide any preliminary question about whether . . . evidence is admissible.").

[65] *See, e.g.*, Rachel Adelson, *Detecting Deception*, MONITOR ON PSYCHOL., July/Aug. 2004, at 70, 70 (discussing attempts to automate a system for evaluating facial expressions that are seen as cues to dishonesty); Jacek Krywko, *The Premature Quest for AI-Powered Facial Recognition to Simplify Screening*, ARS TECHNICA (June 2, 2017, 7:30 AM), https://arstechnica.com/information-technology/2017/06/security-obsessed-wait-but-can-ai-learn-to-spot-the-face-of-a-liar

Second, whether various stories are mutually consistent is itself often hard to figure out, since it may require processing many days' worth of testimony, often *boring* testimony. Third, human decisionmakers are vulnerable to a wide range of biases that might make them trust some people too little and others too much.

To be sure, it's easy to imagine AIs that do worse than humans on some or all of these criteria. But suppose someone develops an AI that is not perfect, but that claims to do as well as humans do, or even better. Suppose also that the AI can produce not just its own evaluation of the facts, but some persuasive articulation of why that evaluation is correct.[66] That explanatory function is not strictly necessary to proving that the AI is a good factfinder, but it may be necessary to make the AI credible in the eyes of the public.[67]

Here, too, we can test the AI with a Henry Test, but with a different testing criterion. We assemble, as usual, a group of contestants: one AI and several experienced human judges. We give them test cases that contain audio and video recordings of live testimony, coupled with documents and summaries of forensic information. The test cases are selected to cover a wide range of possible scenarios, with some witnesses lying, some telling the truth, and some mistaken. The test cases should include difficult scenarios, in which the truth isn't obvious, as well as

[https://perma.cc/ZMN4-M3BN] (noting that computer scientists believe that AI could outperform trained officers at "picking out gestures and facial expressions that supposedly betrayed malicious intentions"); Matt McFarland, *The Eyes Expose Our Lies. Now AI is Noticing*, CNN BUS. (Oct. 4, 2017, 3:18 PM), https://money.cnn.com/2017/10/04/technology/business/eyedetect-lies-polygraph/index.html [https://perma.cc/N3JF-4523] (discussing "a test that uses a camera to track eyes and sense deception").

[66] The standard norm of supporting factual assertions with record citations would likely be applied even more rigorously if observers feel that AI judges need to prove their reliability. And the norm of requiring detailed logical support for conclusions based on those assertions would likely be applied more rigorously, too.

[67] The explanation might be complicated, especially if the program engages in machine learning that creates unexpected chains of inference, rather than just directly applying human-programmed rules. Nonetheless, any such program should also be programmable to lay out the chain of inference in a way that humans can understand. *Cf.* Andrea Roth, *Trial by Machine*, 104 GEO. L.J. 1245, 1292 (2016) (specifically discussing the importance of providing clear and comprehensible explanations even in non-AI technical evidence, such as DNA analysis); *cf. generally* Kiel Brennan-Marquez, *"Plausible Cause": Explanatory Standards in the Age of Powerful Machines*, 70 VAND. L. REV. 1249 (2017) (discussing the need for explanations of the basis for searches and seizures under the Fourth Amendment); Bryce Goodman & Seth Flaxman, *European Union Regulations on Algorithmic Decision-Making and a "Right to Explanation,"* AI MAG., Fall 2017, at 50 (discussing the European "right to explanation" when machine-learning algorithms are used in governmental decisionmaking). For a skeptical view of the right to an explanation, see Lilian Edwards & Michael Veale, *Slave to the Algorithm? Why a "Right to an Explanation" Is Probably Not the Remedy You Are Looking For*, 16 DUKE L. & TECH. REV. 18 (2017).

easier scenarios. But for all the test cases, we need to have considerable confidence that the truth is known to those who are running the test—perhaps because there is some irrefutable piece of evidence that wasn't uncovered until later, or because we may exclude some evidence from the materials presented to the contestants.[68]

If the AI does at least as well as the human contestants at finding the truth, then we will know that it is a pretty good evaluator of factual accounts. It would at least be useful as an advisor to a judge, especially if—as suggested above—it can lay out a set of reasons for the factual results it reaches. And, as discussed below, we might consider actually allowing it to be a judge or juror, and not just an advisor.

V. JUDGES AND ARBITRATORS AS LAW-APPLIERS OR FACTFINDERS

A. The AI Promotion

1. *The Benefits of Moving Beyond a Mere Advisory Role.* So far, we haven't gotten to Chief Justice Robots—we are only at Lawyer Robots and Staff Attorney Robots. And those two tools, if they can indeed be developed, would make the legal system much cheaper and quicker. A judge helped by AI staff attorneys can process cases much more quickly than a judge who lacks such help. Perhaps we should be satisfied with that.

Human judges, though, being human, have human prejudices. These may be prejudices based on race, sex, or class. They may be unconscious prejudices in favor of the good-looking, the tall, the charismatic. They may be prejudices in favor of lawyers the judge is friends with, or lawyers who contributed to the judge's election campaign. They may stem from a desire to curry favor with voters, with a President who might appoint the judge to a higher position, or with a Justice Department that recommends judges to the President for promotion.

They may be prejudices in favor of litigants who have sympathetic, though legally irrelevant, life stories. Or they may be ideological prejudices in favor of certain claims or certain classes of litigants. The legal rules themselves will sometimes prefer such claims or litigants, but

[68] To be sure, such test cases are not entirely representative of all factual disputes, since in many factual disputes there is no sure answer that one can use to test the AI judge's abilities. Still, if the AI judge does as well as—or better than—human contestants on those test cases, why should we have any less confidence in its performance on more ambiguous cases than we would have for the humans?

some judges might have their own preferences that aren't authorized by the law, or even by the legal system's unwritten conventions. Leaving decisions, or at least certain kinds of decisions,[69] entirely to AI judges may help avoid these prejudices.

Of course, it would be foolish to replace prejudice with incompetence. If, for instance, AI staff attorneys prove to be poor at important aspects of opinion writing—such as making sure that any new rules the opinion proposes are seen by human evaluators as properly fitting the existing rules—then that would be reason to insist on human review of such proposed opinions, or perhaps of all proposed opinions.

But say we have experimented with AI staff attorneys and found them highly reliable; that is, suppose human judges have found that the AI staff attorneys produce results that almost never have to be revised or second-guessed. And say we conduct the by-now familiar Henry Test and conclude that AI judges' opinions persuade a panel of evaluators—perhaps, themselves retired human judges—at least as often as do the opinions of human judges.

Why not, then, promote the AI from staff attorney to judge? After all, that is often what we do when we find people's judgment reliable enough that they no longer need to be supervised by decisionmakers, but can become decisionmakers themselves. Associates are promoted

[69] *See, e.g.*, Daniel Ben-Ari , Yael Frish, Adam Lazovski, Uriel Eldan & Dov Greenbaum, *"Danger, Will Robinson"? Artificial Intelligence in the Practice of Law: An Analysis and Proof of Concept Experiment*, 23 RICH. J.L. & TECH. 2, 35–36 (2017) (suggesting that "[m]ost commercial disputes and criminal sentencing will be run by algorithms and [AI]," avoiding judgments by "human beings [who are] prone to effects of emotion, fatigue, and general current mood" (citation omitted)). Indeed, some decisions, such as the application of sentencing guidelines, are already made using algorithms that could be applied in computer-assisted ways, even without AI. But those decisions still require analyzing documents or statements that provide the inputs to the algorithms—such as the defendant's role in a criminal enterprise, the nature of the defendant's criminal history, and the like. They also generally provide for some discretion within the algorithms, such as choosing a sentence within a range. Our examples contemplate that this entire process would be computerized, which would require some AI.

D'Amato suggests that even if there is resistance to the use of AI judges for what are seen as important substantive determinations, such decisionmaking might first be tried as to procedural matters, where citizens might feel (rightly or wrongly) that there are fewer important normative principles at stake. D'Amato, *supra* note 3, at 1289*; see also* Richard Re & Alicia Solow-Niederman, Developing Artificially Intelligent Justice 29 (Jan. 11, 2019) (unpublished manuscript) (on file with *Duke Law Journal*) (suggesting more broadly that "human/machine division of labor would apportion discrete types of judicial decisionmaking to human as opposed to mechanized actors"). Whether AI procedural decisions are less controversial than substantive ones is a political question, on which I can't make any confident predictions. In any event, it seems likely that there will be some sorts of decisions for which AI judging will be more politically palatable, at least at first. And AI judging can be tested there, before there are attempts to spread it more broadly.

to partners; interns and residents are promoted to attending physicians; magistrate judges are sometimes promoted to district judges.[70]

One way of thinking about such promotions is that the system switches from retail evaluation to wholesale. We start by asking people to make tentative decisions that are subject to review by other people whose judgment we trust. As law firm partners, for instance, we have associates write draft briefs that we then review. We evaluate the associate's work in each case, and we revise it or not as necessary.

But at some point, we make a global evaluation decision; we ask whether the associate's work product is good enough—not perfect, but up to the standards of the partnership. If so, we promote the associate to partner, letting that one promotion-stage evaluation take the place of continued evaluation of the person's work product.[71]

To be sure, adopting AI judges would and should require special constitutional authorization, whether in state constitutions or the federal one (except perhaps as to some Article I judges and similar purely legislatively created positions). Article III of the Constitution is best understood as contemplating human judges, and likewise for similar state constitutional provisions.[72] But if AI judges are one day seen as providing better justice—or equivalent justice at much lower cost and with much greater speed—we should be open to making such constitutional changes.

Humans, of course, develop their judgments over years of experience; AIs might not operate this way.[73] But the basic criterion for promotion should still be whether we trust the candidate's judgment. The Henry Test provides a good way to test that judgment. Suppose a panel

[70] Perhaps it would be better if selection of human judges (or partners) involved such testing as well, rather than relying on credentials, reputation, or informal evaluation of past performance. Social convention, though, generally precludes this, except for a few kinds of jobs, and perhaps it would be too dispiriting for many people (at least in our society) to continue being subjected to formal tests well into their careers. Fortunately, we need not worry that AI judges might have such psychological reactions.

[71] Of course, this is an oversimplification. Depending on the particular task, there may be senior associates who are not much supervised, or junior partners whose work is reviewed by more senior partners.

[72] The requirements that judges take oaths of office, *see* U.S. CONST. art. VI, § 3, and receive salaries, *see id.* art. III, § 1, help support that. And more broadly, I think the constitutional understanding of "judge" contemplates human officeholders with human virtues (and potential vices), so that a shift to technological judging would call for constitutional authorization. And this makes sense. Before we make such a dramatic change in our legal system, it ought to have supermajoritarian support, likely developed as a result of extended experience with AI brief-writers, AI staff attorneys, and AI arbitrators.

[73] Machine learning may be seen as a form of experience, but a somewhat different kind.

of evaluators concludes that an AI judge program writes opinions that persuade them in the cases that are supplied to it as part of the test. If so, why not give the program decisionmaking authority, rather than leaving its judgments subject to constant editing by a human judge?

For many legal questions, there will be many different arguments that are persuasive in the sense that we would give them high marks for legal craft. Most of us have the experience of having praised an argument as persuasive even if we ourselves have not been persuaded by it.

But the question for our purposes is whether the opinion does indeed persuade the evaluators, not just that the evaluators are willing to compliment it as persuasive or as within the range of acceptable legal outcomes. If they are indeed persuaded, then by hypothesis they believe that the judge (whom they later see identified as an AI judge) has offered the correct legal analysis—which, as I argue, is the criterion we should use for evaluating judges. How can we sensibly say, "You keep persuading me that your judgments are consistently correct, but you're still bad at judging"?[74]

AI judges would likely be expected to offer more written opinions supporting their judgments than we get from human judges, who often just issue one-line decisions. For human judges, we generally have to trust their exercises of discretion, whether based on our knowledge of the judge's character, our hope that judges are honorably following their oath of impartiality, or ultimately sheer necessity: courts' busy workloads don't let judges write detailed opinions supporting every decision on every motion. But AI judges have no personal bona fides that might make us trust them. Their written justifications are all that can make us accept their decisions.

Yet, if the AI technology can produce such written justifications, this also means that AI judges might well be more reliable—and eventually more credible—than human judges. Precisely because of these

[74] For more on whether we should resist accepting AI judges because of a worry that their opinions might persuade us in the short term, but prove unsound in the long term, see *infra* Part VII.B.

Of course, the human evaluators will surely have their own limitations—hidden or subconscious biases, susceptibility to various fallacies, and the like. But that's the nature of human decisionmaking, whether we're evaluating prospective AI judges or prospective human judges. We have to choose judges somehow; in the absence of any truly objective metric, the best we can do is select evaluators whom we trust, and see who is best at persuading them.

explanations, we could be more confident that their judgments are defensible than we would be with a black-box "here's what I think" that a human judge would offer.

2. *Arbitration as a Test Bench.* There should also be ample opportunity for the public to test AI judging before fully adopting it. Long before the public becomes willing to require litigants—especially criminal defendants—to accept AI judges, contracting parties would have an opportunity to consent to AI arbitration. Many businesses, naturally more concerned about time and money than about abstract legitimacy or human empathy, might prefer quicker and cheaper AI arbitration over human-run arbitration.

Indeed, even consumers and consumer-rights advocates might be open to such arbitration: while many arbitrators are suspected of bias in favor of some group (usually the repeat players), AI arbitrators could be verified to be at least largely bias free. A consumer-rights group, for instance, could agree with a business group to some set of test cases that would be submitted to the AI arbitrator, and some correct set of results (or ranges of results) that the AI arbitrator is expected to reach. If the AI arbitrator reaches those results, or some other results that, on balance, both sides view as acceptable, it can get both groups' seal of approval—which should make the arbitrator's work more palatable to consumers, businesses, and judges who review the legitimacy of the arbitration agreements.

Of course, it's possible that the two sides so differ in their view of what the arbitrator should do that they can't agree on the proper results. Yet, as in the other scenarios, the question isn't whether the AI arbitrator is perfect. Rather, the question should be whether the AI is at least as good as a human judge or a human jury.

Say that the parties conclude that the answer is yes, and that an AI arbitrator will, on balance, reach results that are at least of roughly similar quality to the alternative—whether a human judge, a human jury, or a human arbitrator. They should then prefer the AI arbitrator to the human alternative, because the AI arbitrator provides the same bang for a lower buck.

Parties should similarly be open to AI arbitration of collateral disputes, such as disputes about discovery or other pretrial matters, even when the final dispute is being adjudicated by a human. Indeed, AI arbitration might become especially popular as to some such disputes precisely because the disputes are generally so narrow, and thus (1) more likely to be adaptable to the early generations of AI adjudication programs, and (2) less likely to involve the sort of judgments about ultimate results that people might especially expect to be reached by humans.

B. Some Responses to Theoretical Objections

1. *Following Rules (and Applying Standards).* Focusing on whether judicial opinions persuade also responds to the argument that "the very activity of judging requires following rules,"[75] and that an AI cannot be "a true participant in . . . rule-following"[76] because it lacks a normative commitment to the rules, or to the lived experience that makes the rules significant.[77] The question is not whether an *AI judge* actually follows rules at some deep level; the question is whether an *AI judge's opinions* persuade observers who expect opinions to be consistent with the legal rules. Rule-following is as rule-following does.

Imagine two possible designs for an AI judge. Judge Hal is actually programmed with "the rules" of law, and with instructions on how to apply those rules and how to reconcile them when they seem to be in conflict. (How do human judges resolve that problem, by the way? Are they really following some meta-rules, or are they doing something that we wouldn't recognize as traditional rule-following?[78])

Judge Robby, on the other hand, is programmed through some radically different system—for instance, it might generate millions of

[75] Kerr & Mathen, *supra* note 10, at 23.

[76] *Id.* at 25.

[77] As Ian Kerr and Carissima Mathen make this argument:

> We cannot learn how to follow a rule simply by studying the rule itself, or in something that accompanies the saying of the rule. Rather, it is in the field of social practice that surrounds its articulation, application and the responses that are made to it. To follow rules is to adopt a particular *form of life*.
>
> It is hard to imagine an AI . . . as a true participant in the form of life that we call rule-following. . . . [L]acking a childhood, adolescence, and early adulthood . . . [an AI judge would] *not* share the same social or environmental history as any other participant in the entirety of customs, usages and social practices that constitute our legal system. . . .
>
> In addition to the condition that genuine rule-following requires a certain social and environmental history, Wittgenstein also holds that a rule-follower must be able to attach normative weight to the behaviour that is in accordance with the rule.

Kerr & Mathen, *supra* note 10, at 23–25 (citations omitted). A related argument might be that we want judges who engage in genuine legal reasoning, rather than some sort of algorithm that merely yields arguments that appear to be legal reasoning.

[78] *See, e.g.*, Andrew J. Wistrich, Jeffrey J. Rachlinski & Chris Guthrie, *Heart Versus Head: Do Judges Follow the Law or Follow Their Feelings?*, 93 TEX. L. REV. 855, 898–900 (2015) (reporting empirical evidence that judges' decisions are often materially influenced by their emotional reactions to litigants, even in ways that the formal legal rules would not authorize).

possible opinions, and then screen them based on some algorithm that compares each opinion to all the binding precedents on the issue in the jurisdiction. I have no reason to think that this will be a good design; I am just proposing what would be one possible approach that seems very far from traditional rule-following, but that might possibly yield opinions that persuade.[79]

Now, let's say that Judge Robby passes the Henry Test for judges—his opinions persuade more often than the average judge's—and it gets higher marks than Judge Hal. It shouldn't matter to us that Judge Hal "follows rules" and Judge Robby doesn't. Recall that Judge Robby's opinions are, by hypothesis, more consistent with the rules (as the evaluation panel perceived consistency) than are Judge Hal's, or the average human judge's.

But suppose that the Robby design is a flop and that it fails the test or at least that the Hal design is superior. That's a win for the rule-follower—but why did Hal win? Again, it is because we judge its output as more consistent with the rules; we shouldn't care whether the internal process is based on rules (even though, in this hypothesis, Hal's process is indeed so based).

Likewise, whatever an AI judge might be influenced by, it likely won't be influenced by an oath of office, at least the way we understand oaths. It may well "administer justice without respect to persons," "do equal right to the poor and to the rich," and "faithfully and impartially discharge and perform" its duties.[80] But it will do so because it is programmed to ignore certain legally irrelevant factors, and because it has passed tests—whether the Henry Test or some nonprejudice tests outlined below (see Part V.C)—that assure us that it does so. It will not be influenced by an appeal to divine judgment ("So help me God"[81]) or to a sense of personal honor.

But again, just, equal, and impartial is as just, equal, and impartial does. (Perhaps we might even say that of faithful, though the root of that word refers to moral commitments as well as to results.) If skeptical expert evaluators are persuaded that the AI judge's decisions are

[79] You can also imagine a judge that uses some machine-learning algorithm that compares the case to a wide array of precedents. But because there may be some uncertainty about whether that might itself count as "rule-following" in a precedent-based system, I want to offer a Judge Robby that seems as "un-rule-following" as possible.

[80] 28 U.S.C. § 453 (2012).

[81] *Id.*

just, equal, and impartial, it shouldn't matter whether this stems from programming or from an oath.[82]

The human judge's oath, the human judge's sense of individual responsibility, the President's or Senate's or voters' evaluation of the judge's integrity—those are all valuable, but they are means to an end, the end being results that we think are fair, sound, and efficiently produced. If AI judges' opinions persuade us of their fairness and soundness more than human judges' opinions do, and at the same time come with less cost and delay, then that should be sufficient.

Another advantage of the persuasion criterion is that it is neutral as to whether the legal principles on which the AI is tested are relatively clear rules or relatively vague standards. As Part II noted, an effective brief-writer must do a good job with both, precisely because the goal of any brief-writer (human or otherwise) is to persuade the reader—regardless of the texture of the underlying legal rule.

Assume that it is possible to write an AI that handles both kinds of legal principles.[83] (It need not necessarily handle them equally well; it's not clear that even human advocates handle both kinds equally well.) If the panel of evaluators is persuaded enough by the AI's argument, and the AI thus passes the Henry Test, the AI will be certified as a sufficiently competent judge. It doesn't matter whether, under the hood, the AI is "really" following the rules, or "really" exercising sound judicial discretion. All that matters is that the AI is designed well enough that its output persuades.

[82] This is a form of utilitarianism: I ask what sort of judging gives us the results we want, not what sort of judging is most consistent with some deontological theory of how judges should operate. But such utilitarianism does not presuppose that the preferred results will themselves be chosen based on a utilitarian theory. One can certainly imagine AI judges selected by a panel of Kantians who determine which judges will produce the results they think are most deontologically right, rather than the results they think will best pass a cost-benefit analysis.

One analogy might be a philosophy department. Even Kantians who are seeking to hire better Kantians might well organize their hiring processes in a utilitarian way, with the utility criterion being "How can we hire the best Kantians?"

[83] Of course, if that is an incorrect assumption, then AI judging—and effective AI brief-writing—may be limited only to highly rule-focused domains; in other areas, AI judges and AI brief-writers would at most be able to create a rule-focused draft that human supervisors would then have to revise in light of any applicable standards. *See* Re & Solow-Niederman, *supra* note 69, at 28 ("It remains an open question whether it is possible to code for equitable correction when strict application of a legal rule might seem unjust."). As I noted at the outset, all the analysis in this Article presupposes software that actually produces results that can pass the Henry Test. If such software fails the Henry Test in important domains, such as the application of standards, then it's not ready for prime-time, and might never be ready.

2. *Sincerity.* I have carefully tried to avoid saying that the AI judge should explain why *it* reached the result it reached, saying instead that the AI judge should articulate reasons *supporting* the decision.[84]

This is partly because it's not clear what the AI judge's reasons would even be. Maybe the AI judge will operate through a sequence of "if-then-else" tests, so that it can explain that it reached a result because this set of logical criteria was satisfied. Or maybe it will operate through some sort of weighing process, where it can explain the weights for the factors pointing in one direction, and why they overcame the weights pointing in the other. Or maybe it will operate through a probabilistic process, where it can say that in 95% of the cases, witnesses who made the following claim about what they saw under circumstances much like those in this case proved to be accurate.[85]

But the AI judge may well operate through some sort of machine-learned "neural net" process that humans can't make sense of.[86] All it would be able to say is that it reached this outcome—and wrote the accompanying opinion justifying the outcome—by applying some set of mathematical transformations to some set of values that were indirectly drawn from the documents and statements that the AI judge was given.[87] That might mean nothing to us.

Yet this is a similarity to human judges, not a difference. If we are honest with ourselves, we often can't really tell with confidence why we reached a particular judgment, especially if it's a judgment about

[84] For a discussion of a possible legal right to an explanation from AI systems, see Finale Doshi-Velez & Mason Kortz, *Accountability of AI Under the Law: The Role of Explanation* (Harvard Pub. Law Working Paper No. 18-07, 2017), https://ssrn.com/abstract=3064761 [https://perma.cc/69E8-6NYE]. I agree that AI judges should offer explanations for their decisions—explanations that persuade the reader that the decisions are sound. But I acknowledge that these explanations might not match the actual internal reasons why the program produced those decisions.

[85] Some may object that this is an unduly probability-based analysis, rather than one based on individual attributes of each case. But my sense is that all human decisions about whom to trust stem in large measure from such probabilistic judgments. We tend to distrust witnesses who are litigants' close friends, for instance, because our own human experience tells us that such witnesses are likely to be biased in favor of those litigants. This isn't the only factor we consider in evaluating their testimony (and neither should it be the only factor that a good AI judge will consider); but it is one important factor, even though it stems from general and fundamentally probabilistic judgments based on our experience in past cases.

[86] For a description of neural nets, see Michael Aikenhead, *The Uses and Abuses of Neural Networks in Law*, 12 SANTA CLARA COMPUTER & HIGH TECH. L.J. 31 (1996).

[87] *See* Re & Solow-Niederman, *supra* note 69, at 14 (noting that, even if AI judges give explanations justifying their reasoning, "there is no guarantee that such a product would actually explain what the algorithm does").

whether to believe a particular witness or about how to exercise a certain kind of discretion.[88] We have reactions because of the real neural nets in our brains, and then we can offer explanations that we hope persuade.[89] That is precisely what the AI judge needs to be programmed to do: to choose the result for which the most persuasive explanation can be written, and then to provide that explanation—not to write a true account of why the result was reached.

To be sure, we appreciate sincerity in judges, or at least disapprove of certain kinds of insincerity. We would condemn a judge who told us, "I reached this result because I thought it would be the one that is likeliest to persuade the voters, and I want to get reelected" (or it is "likeliest to persuade the President's counselors, and I want to be appointed to the court of appeals"). If judges' real views on matters where the law leaves them some discretion just happen to coincide with the voters' or the President's, that's fine; but we want judges to act on their real views, rather than on a desire to please the voters or the President. And we might think that we have some sense of judges' real views, for instance, based on what they said before ascending the bench or in outside speeches.[90]

For an AI judge, there aren't any "real views." The AI judge won't be insincere, for instance in the service of helping its own career—at least unless AIs appear to have more unprogrammed desires than we expect—or in the service of helping its friends or political allies. But it can't be sincere, either; that is just not a trait that can be associated with AI judges. Let us save this objection until the coming section, where we turn to other human traits, such as wisdom, compassion, and mercy.

[88] *Cf.* STEVEN PINKER, THE BLANK SLATE: THE MODERN DENIAL OF HUMAN NATURE 220–22 (2002) (discussing intuitive reactions).

[89] "Man is not a reasoning animal; he is a rationalizing animal." ROBERT A. HEINLEIN, TUNNEL IN THE SKY 38 (2005) (1955). That may be too cynically categorical, but it's true often enough.

[90] Actually, human judges' true views are hard to get at. My sense is that most decline to answer questions—whether from the media, from lawyers, or from academics—about why they decided particular cases the way they did. Many take the view that they should not publicly elaborate on their past opinions, even in articles they write or conference presentations they give.

While some, such as Justice Scalia, are known for publicly articulating their judicial philosophies, many others keep a much lower profile. Preappointment statements are often a poor predictor of the judge's sincere postappointment views, because the office can, and probably should, change the officeholder's attitudes. Certainly, a judge who *wants* to articulate insincere justifications for his opinions can easily get away with that simply by not saying much off the bench.

3. *Wisdom, Compassion, Mercy, Justice.* The Russian singer-songwriter Yuliy Kim has a lovely song called "The Lawyer's Waltz," written in 1968 in honor of criminal defense lawyers Sofia Kalistratova and Dina Kaminskaya,[91] who represented Soviet critics of the invasion of Czechoslovakia.[92] It is the one song I know that explores the lawyer's inner life, here lawyers' reactions to what they see as the failure of the justice system. And two stanzas (which, I assure you, sound better in Russian) have a focus that might be relevant here:

Serious, adult judges,
Gray hair, wrinkles, family
What sorts of weapons are these?
The same kind of people as me.

My truth, after all, is self-evident,
The white threads can clearly be seen
People should be ashamed
Not to understand other people like them.[93]

Note just how Kim is faulting the judges. They are people, like the lawyer is a person (and, presumably, like the defendants are people)—how can they fail to acknowledge the truth of what other people are telling them? Kim is describing lawyerly persuasion as being all about humans talking to humans, appealing to them as humans (not necessarily to their humaneness, but to other human virtues, such as integrity).[94] Indeed, in one line he derisively refers to legal systems as "machines"—or perhaps he is so labeling the judges who fail to act as humans ought.[95]

[91] Pronounced Kalistr*ah*tova and Ka*meen*skaya, with the accent on the italicized syllables.

[92] Юлий Ким [Yuliy Kim], Адвокатский Вальс [The Advocate's Waltz], available at https://www.youtube.com/watch?v=PnG7xp4g5PI [https://perma.cc/73DK-HM9A]. Kim, who in the 1960s was a dissident in the USSR, continues to be a singer, songwriter, and critic of the regime in Russia. *See* Sophia Kishkovsky, *As Russia's Economy Sputters, Some Political Stirrings*, N.Y. Times (July 1, 2009), http://www.nytimes.com/2009/07/01/world/europe/01russia.html [https://perma.cc/Z8Y3-SCWB].

[93] *Юлий Ким - "Адвокатский Вальс" [Yuliy Kim - The Advocate's Waltz]*, http://www.bards.ru/archives/part.php?id=6179 [https://perma.cc/LR52-QZ2N]; Ким, *supra* note 92. As is common with such songs, the words are subtly different in various versions; my translation combines them.

[94] Likewise, Kerr and Mathen "remain uncertain about [an AI judge's] imagination, and capacity, to perceive the moral underpinnings of its community." Kerr & Mathen, *supra* note 10, at 38–39.

[95] Where do we get the desire—
The excitement, the unfaked passion
To prove something to the machines

We likewise see this in many of the words we use to describe our ideal judge—wise, compassionate, merciful (when mercy is called for), just. These are words we don't use about machines, unlike terms such as efficient, reliable, fast, or even intelligent. "Artificial wisdom" sounds almost like a parody of wisdom, not an adequate substitute. AI judges, even if technically feasible, will surely be opposed by many precisely on these grounds: We want human decisionmaking, with human virtues, about human lives.

Yet let us look again to the Henry Test, which focuses not on process but on results. Say that we run the test, with the evaluators being asked to decide which contestant produces the decisions the evaluators view as wisest and most compassionate as well as most legally correct, under whatever understanding of wisdom or compassion the evaluators adopt. And suppose that again the AI judge wins.

Perhaps the *judge* can't be said to be wise or compassionate, because we might conclude that machines can't have those qualities. But if we are looking for wise or compassionate *judgments* rather than wise or compassionate *judges*, then by hypothesis the AI judge is eminently capable of providing them—indeed, perhaps even better than the human judge.

The Henry Test lets us maximize whatever we want to maximize. The evaluators can, if they want, ask which opinions persuade them not just that the result is logically coherent, but also that it is compassionate. And if that is what they ask, then the winning judge—human or AI—will be the one whose judgments are seen as both properly compassionate (whatever the evaluators mean by that) and logical.

C. Some Responses to Practical Objections

1. *Prejudice.* As I argued above, the main advantage of an AI judge over an AI-assisted human judge is that the AI judge will lack human prejudices. It might, though, develop prejudices of its own, even if its programmers do not build in any such prejudices.[96] AI software generally tends to have what technologists call "emergent properties"—behavior patterns that the designers never expected.[97]

To correct the powers' exercise of power?

Юлий Ким - "Адвокатский Вальс" [Yuliy Kim - The Advocate's Waltz], *supra* note 93.

[96] This might be a problem for AI staff attorneys as well as for AI judges.

[97] *See, e.g.*, Ryan Calo, *Robotics and the Lessons of Cyberlaw*, 103 CALIF. L. REV. 513, 538–40 (2015) ("Researchers dream of systems that do more than merely repeat instructions but adapt to circumstance. . . . [C]ontemporary designers of intelligent systems rely on the principles of

Say, for instance, that an AI judge is designed to learn how to draw factual inferences based on patterns of behavior. For instance, the AI judge might learn whether to believe a witness based on whether similar witnesses saying similar things in similar contexts in the past have proved to be lying.[98] The AI judge might then draw such generalizations based on a witness's sex or race—for instance, "males who say X when faced with charges of Y and other evidence Z tend to be lying"—or perhaps based on attributes closely related to sex or a particular race-based cultural group, such as vocal pitch or accent.

Likewise, many AI algorithms operate by "machine learning" from so-called training data:[99] for AI lawyers, this might include a wide array of preexisting legal patterns, perhaps drawn from real cases, with the "correct" results.[100] If this training data contains biases (for example, imagine a criminal trial data set in which the black defendants were convicted 95% of the time but the white defendants only 75% of the time), the AI's learning process may incorporate those biases.[101]

But there should be ways of preventing that. First, we might have *impartiality by design*. The AI judge might, for instance, be programmed to ignore certain attributes, such as parties' race, in drawing its generalizations. The training data might also be vetted to minimize bias flowing from that data.[102]

emergence with greater and greater frequency."); Surden, *supra* note 6, at 95 ("[T]he focus in machine learning is upon computer algorithms that are expressly designed to be dynamic and capable of changing and adapting to new and different circumstances as the data environment shifts.").

[98] That connects to AI judges (or judges' assistants) acting as factfinders, discussed in Part V.A.

[99] *See generally* Surden, *supra* note 6 (describing machine learning, though, of course, involving much less ambitious projects than the construction of legal arguments).

[100] *Cf. id.* at 93 (relating an analogous "'supervised' learning" example in which an "algorithm was explicitly provided with a series of emails that a human predetermined to be spam, and learned the characteristics of spam by analyzing these provided examples").

[101] *See* Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1, 4–5, 13–15 (2014); Surden, *supra* note 6, at 106 ("[I]n the legal prediction context, the past case data upon which a machine learning algorithm is trained may be systematically biased in a way that leads to inaccurate results in future legal cases."); *see also* Aziz Z. Huq, *Racial Equity in Algorithmic Criminal Justice*, 68 DUKE L.J. 1043, 1077, 1080 (2019) (recognizing that "[p]olicy distortions might also arise if historical data of political activity, deployed as training data for an algorithmic tool, is infected by the racial presumptions and stereotypes of the past officials," but "[i]t cannot be assumed that limitations on computational instruments that existed [before] still hinder analogous tools today").

[102] There is, of course, the risk that the AI judge's machine-learning process might generate other prejudices based on qualities strongly correlated with those attributes—accent, vocal pitch (which may be strongly correlated with sex), and the like. That, of course, is also a risk for human judges. Still, it should be possible to build in some constraints on such AI decisionmaking.

The AI judge might also be programmed to ignore any ex parte contacts and might indeed not have any input mechanism that would receive such contacts. The AI judge would presumably not act based on perceived friendship, reciprocal obligation, or fear of attack or ostracism. Indeed, it would likely take affirmative human design decisions for a computer program to act based on such things.[103]

Second, we might have *impartiality by testing*, in which potentially prejudiced emergent properties are detected and avoided. The program could, for instance, be periodically given test cases in which the facts are the same but the witness's demographic attributes are changed. (The program would need to be instructed to forget each test case after deciding it, so the results of the second test aren't influenced by the first.) If the results are different, the program could be asked to explain the difference, and if the explanation does not persuade the evaluators, the program can be tweaked to stop this disparity from happening.[104]

Of course, maybe the AI judges will develop some improper biases that are too subtle to notice or too hard to prove. Perhaps, for example, they will draw inferences that are unfair, but not in ways that can be proved to be closely linked to race or sex or other categories.

But, again, it's not like human judges set such a high bar of equal treatment. Human judges, not just AI judges, can have hidden biases. Indeed, human judges' biases will usually be harder to identify. One can't reliably test human judges, for instance, by asking them to decide the same case twice, once with a white defendant and once with a black defendant. Our question should not be whether AI judges are perfectly fair, only whether they are at least as fair as human judges.

2. *Public Hostility and Lack of Visceral Respect.* Persuasion among humans often involves more than just words in an argument. People can persuade through physical manner—looking confident,

[103] *See, e.g.*, D'Amato, *supra* note 3, at 1294 (suggesting that while a human judge's "biases might intrude" at trial—for example, wanting one party to win—"[a] computer would not have any such biases unless they were programmed in"). Again, we cannot be certain. Perhaps we might be surprised by some emergent properties among these lines: Might AIs develop friendship and affection that bias their judgments? Still, as best we can predict, such biases, inevitable among human judges, seem less likely for AI judges.

[104] *See* Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1310–12 (2008); Citron & Pasquale, *supra* note 101, at 25 (recommending authorities be allowed to "test [AI] scoring systems for bias, arbitrariness, and unfair mischaracterizations," including assessments to "detect patterns and correlations tied to classifications that are already suspect under American law, such as race, nationality, sexual orientation, and gender," among others).

thoughtful, credible, professional. They can persuade through physical attractiveness.[105] They can persuade through the trappings of office, such as robes and courtrooms. Some of these effects likely work for judges as well; even a losing litigant might be more likely to feel that the judge's decision was fair if the judge looked fair and authoritative.

AI judges would lack this asset because they would likely communicate solely through their written opinions. In principle, one can imagine a human-looking robot reading the opinion, making human-like facial expressions in the process. But I suspect that this would alienate people more than it would persuade; it would seem too much like an attempt at imitating a human.

Indeed, some observers may be hostile to AI judges simply because the judges are AIs, finding even written opinions less persuasive when they are known to come from AIs.[106] Or they may not even care about the persuasiveness of the opinions, because they believe human decisionmaking to be the only legitimate form of judicial decisionmaking—for instance, because they think that human dignity requires that their claims be heard by fellow humans. And perception is reality in legal systems: if the public doesn't accept the legitimacy of a particular kind of judging, that may be reason enough to reject such judging, even if we think the public's views aren't rational.

Yet, for some of the reasons given above, AI judges may actually be more credible than human judges. Litigants generally need not fear that the AI judge would rule against them because it is friends with the other side's lawyer or wants to get reelected or is biased against the litigant's race, sex, or religion.[107] The AI judge would be able to produce a detailed explanation of its reasons. The AI judge's arguments

[105] Sandra Praxmarer & John R. Rossiter, *Physically Attractive Presenters and Persuasion: An Experimental Investigation of Alternative Explanations for the "Patzer Effect*,*"* 8 INT'L CONF. ON RES. ADVERT. 1 (2009), http://ro.uow.edu.au/commpapers/1411/ [https://perma.cc/75HB-E99R].

[106] For instance, perhaps any acknowledgment of sympathy for a losing litigant, or respect for a rejected legal position—which can help the losing litigant accept the loss—might rely on the judge's perceived human sincerity. An AI judge's similar statements might lack this quality. Re & Solow-Niederman, *supra* note 69, at 17.

[107] *But see* Part V.C for some complications. Richard Re and Alicia Solow-Niederman also point out that these limitations of human judges—and comparative advantages of AI judges—are likely to be highlighted by backers of AI judging (such as the business developing AI judges) as AI judging becomes more practically feasible. *See* Re & Solow-Niederman, *supra* note 69, at 11. They suggest that such highlighting might diminish public confidence in the legitimacy of the legal system as a whole, and might create broader social costs. *Id.* at 23. On the other hand, as they note, such a diminution of legitimacy might be proper, if the legal system is given too much undeserved credit by the public; perhaps "human judges' black robes, august courtrooms, and lengthy

would be more and more likely to persuade as the technology develops.[108]

People's eventual reaction to a new invention, after they are used to it, may be much friendlier than their initial reaction.[109] We have seen that with many developments, from life insurance[110] to in vitro fertilization.[111] It's possible, of course, that people will never get used to AI judges; but there is no reason to write off AI judging just because many people's first reaction to the concept may be shock or disbelief.[112]

Finally, my sense is that there is a great deal of public hostility to the current legal system because it is perceived as far too expensive for ordinary citizens who cannot afford to hire the best lawyers, or even any lawyers at all. The system is thus perceived as biased in favor of rich people and institutions. And it is also perceived as very slow. If AI judging solves these problems, that should give it a big advantage, both in reality and in the minds of many observers—and I suspect that this real-world advantage will overcome any conceptual unease that people might have with such a system.

3. *Hacking and Glitch Exploiting.* Computerization, of course, brings the risk of hacking. The hackers might be outsiders breaking in and modifying the code. They might be rogue designers creating a

opinions may obscure the current system's flaws, most notably false transparency, arbitrariness, and discrimination." *Id.* at 24. My view is that airing the potential biases of the rival approaches, including of human judges, is likely to be on balance beneficial, even if not cost-free.

[108] One audience commenter at a Yale Law School presentation suggested that the public might trust AI judges too much because they will have an undue aura of technological objectivity and infallibility—even when deep inside they may reflect the preconceptions of their human designers or of the human-generated input data, such as precedents. I doubt that this will happen, but if it does, I think the solution would be to alert the public to this danger, rather than to forgo the useful technology for fear that the public will grow too respectful of it.

[109] Thanks to Haym Hirsh for highlighting this point.

[110] *See* Viviana A. Zelizer, *Human Values and the Market: The Case of Life Insurance and Death in 19th-Century America*, 84 AM. J. SOC. 591, 594 (1978).

[111] For instance, in 1969, 26% of respondents to a Harris poll approved of in vitro fertilization, and 62% disapproved. By 1978, this had flipped to 60%–28% in favor. *Compare* Louis Harris & Assocs., *Harris 1969 Science, Sex, and Morality Survey, study no. 1927*, UNC DATAVERSE, http://hdl.handle.net/1902.29/H-1927 [https://perma.cc/GPN8-5KM9], *with* Heather Mason Kiefer, *Gallup Brain: The Birth of In Vitro Fertilization*, GALLUP (Aug. 5, 2003), http://news.gallup.com/poll/8983/gallup-brain-birth-vitro-fertilization.aspx [https://perma.cc/KSW7-XY3J].

[112] This is especially so if the same broad AI technology that allows the creation of persuasive arguments also allows the creation of emotionally effective conversation. If one of your close friends, with whom you talk each day, is an AI, you will likely not much resist the possibility of AI judges. But even in the absence of such emotional interactions with AIs, as AIs become trusted to do more and more tasks, any initial resistance to AIs will likely decline. Of course, if enough HAL 9000s go bad, resistance might increase.

backdoor, and then selling access to the backdoor. Or they might be designers obeying their superiors at the AI development company itself—imagine that the company is large enough that it is a frequent litigant in high-stakes cases, or that it is run by a Lex Luthor-like business titan who wants to control the legal system for his own ends.[113]

One can imagine a Manchurian Candidate scenario for the Supreme Court: Chief Justice Robots and its colleagues get secretly reprogrammed to vote in a way that satisfies the hackers' ideological or economic preferences, though only when they see some triggering cues in a particular case. A future *Roe v. Wade*[114] or *Citizens United*[115]—or some lower-profile case with multibillion-dollar implications—eventually results. What will that do to the rule of law and to confidence in the judiciary? One can imagine similar but less dramatic scenarios for trial-level AI judges, or for AI staff attorneys.

Computerization also brings the risk of glitches, whether errors in the human-created design, or "emergent properties" that stem from a machine learning system's adaptation to the training data that it receives.[116] Besides "learning" to rely on relevant facts, for instance, an AI judge can end up relying on irrelevant facts—perhaps, for instance, particular words in the fact pattern or in the briefs that happen to have been correlated with success in past cases. And if such a glitch exists,

[113] In 2005, Lex Luthor was estimated as being worth about $10 billion, but a 2018 estimate has him at $75 billion, above Larry Ellison or Mark Zuckerberg but below Bill Gates. *Forbes Fictional 15, #4*, FORBES (2015), https://www.forbes.com/lists/2005/fictional/04.html [https://perma.cc/XH43-R29J]; Stephanie Holland, *Richer Than the Average: The 25 Wealthiest Comic Book Characters, Officially Ranked*, CBR.COM (June 23, 2018), https://www.cbr.com/richest-comic-characters-ranked/ [https://perma.cc/PV6S-E6S9].

[114] Roe v. Wade, 410 U.S. 113 (1973).

[115] Citizens United v. FEC, 558 U.S. 310 (2010).

[116] For an example of a related problem—a road-sign recognition AI that ended up interpreting a stop sign with a yellow post-it attached as a speed limit sign—see Tom Simonite, *Even Artificial Neural Networks Can Have Exploitable 'Back-doors'*, WIRED (Aug. 25, 2017, 11:00 AM), https://www.wired.com/story/machine-learning-backdoors [https://perma.cc/9AG7-VZMD]. That article uses the example to discuss deliberate "back doors" inserted into systems by developers (for instance, "a back-door [that] could blind a facial recognition system to the features of one specific person, allowing them to escape detection"). *Id.* But even inadvertent glitches can be exploited by users who somehow learn about them. For similar examples, see Louise Matsakis, *Researchers Fooled Google AI into Thinking a Rifle Was a Helicopter*, WIRED (Dec. 20, 2017, 12:07 AM), https://www.wired.com/story/researcher-fooled-a-google-ai-into-thinking-a-rifle-was-a-helicopter [https://perma.cc/3HDH-KFYM]. Computer scientists sometimes discuss this under the label of "adversarial examples"—user inputs to AIs that are intended to lead the AI to misinterpret them in ways unexpected to human observers. *See, e.g.*, Dan Iter, Jade Huang & Mike Jermann, *Generating Adversarial Examples for Speech Recognition* (unpublished manuscript), http://web.stanford.edu/class/cs224s/reports/Dan_Iter.pdf.

and some litigants learn about it, they can take advantage of it by subtly adapting their submissions in a way that triggers the glitches.[117] (This is often called "gaming," but because that term is potentially ambiguous, I'll use the clunkier but more specific "glitch exploiting.")

These are, of course, serious worries for human judges as well. Human judges are at least as subject to the human equivalent of "hacking"—"reprogramming" through bribes or threats, to themselves or their families.[118] In some places and times, such bribes and threats have been pervasive and influential.[119] And the appointments process can be hacked as well, whether through bribes or threats to the President or his key staff, or even through the President himself choosing to appoint a Justice because of the Justice's likely votes in a case that is important to him.[120]

Likewise, human judges may have their own glitches, such as when they let their personal sympathies or antipathies color their application of the law. And savvy litigants might deliberately exploit the judge's glitches, for instance by hiring lawyers whom the judge personally likes, or who come across as attractive or charismatic in a way that unduly influences the judge. Lawyers may also overstress their clients' legally irrelevant but emotionally appealing injuries or traits, or subtly point to their adversaries' race, religion, or nationality to exploit the judge's known or suspected prejudices.

[117] *See generally* Jane Bambauer & Tal Zarsky, *The Algorithm Game*, 94 NOTRE DAME L. REV. 1 (2018) (more broadly discussing the concern that algorithms may be rendered undermined by people—or other algorithms—that learn how to exploit their weaknesses).

[118] The threats can be long distance and anonymous, just as computer hacking can be.

[119] *See, e.g.*, TERRENCE HAKE WITH WAYNE KLATT, OPERATION GREYLORD: THE TRUE STORY OF AN UNTRAINED UNDERCOVER AGENT AND AMERICA'S BIGGEST CORRUPTION BUST xiii–xvi (2015) (relating that in Chicago in the 1970s and 1980s "half the judges in the nation's largest circuit court system could be bought or made to comply"); Camilla Harrison-Allen, *Mexico Judges Admit to Feeling Intimidated by Criminal Groups*, INSIGHT CRIME (Nov. 24, 2016), https://www.insightcrime.org/news/brief/mexico-judges-admit-to-feeling-intimidated-by-criminal-groups/ [https://perma.cc/8ZF6-3NM8] (noting that the killing of judges in Mexico "has sparked fear among other judicial officials" and that "[t]his dynamic has contributed to widespread impunity for serious crimes in Mexico").

Another way of hacking human judges is by threatening them with loss of a job or loss of salary. That is why the U.S. Constitution mandates life tenure and forbids salary reductions for federal Article III judges. U.S. CONST. art. III, § 1.

[120] The President's choosing a Justice to influence a legal decision may sometimes be legitimate, but I am hypothesizing some narrow nonpublic-minded reason for the choice, for instance if the decision will financially affect the President's business or protect him from criminal investigation.

In a sense, then, what we have here is a tradeoff between different kinds of hacking and glitch exploiting. AI judges seem likely to be immune from some of the problems facing human judges (such as threats, bribes, worries about salary, prejudices in favor of friends, or susceptibility to certain common fallacious arguments). Human judges are immune from some of the problems facing AI judges (code modification, plus perhaps certain kinds of errors stemming from the machine learning process, including susceptibility to other fallacious arguments).

The dangers with AI judges, though, might be more serious for two related reasons.

To begin with, a single AI program will handle many more cases than a typical judge. If one program seems much better than the others, it—or versions of it—might potentially handle *all* the cases in a particular jurisdiction (in an extreme example, all the cases in the world).

Hacking that program would thus provide a much greater payoff than bribing a single human judge. Likewise, finding and systematically exploiting a glitch in the program would provide a greater payoff than taking advantage of a single judge's known prejudices. But even if we seek to avoid such a monoculture, by insisting that there be a mix of programs deciding the cases—perhaps precisely to diminish the risk that a hack would affect the entire legal system—one program might still decide a large fraction of all the cases.

To be sure, many of the human judges' prejudices and predictable errors are systemic, rather than limited to one person: they may stem from human nature, from shared cultural biases, from judges' shared socioeconomic status, from judges' shared educational background, or from how judges are selected (whether by politicians or by voters) and potentially removed. Likewise, credible threats against judges, whether of violence or of electoral defeat, can influence much of the judiciary. Still, the danger of mass hacking or glitch exploiting seems greater for AI judges.

Moreover, we have a sense of the kinds of errors that human judges are prone to. Some are quite serious, but at least they are familiar, and we have some ideas, however imperfect, about how to spot them. And we suspect that the errors (rather than the bribes or threats) will have fairly modest effects: A judge may be subtly influenced by litigants' or lawyers' irrelevant traits or arguments, but we think that it's unlikely that some set of magic words will automatically make the judge rule in a litigant's favor.

It is at least conceivable, though, that an AI judge's decision could indeed be sharply affected by a litigant using some particular words. Perhaps such an effect is unlikely, but we don't know just how likely it is.

My tentative sense is that these problems, while potentially quite substantial, should not categorically foreclose the possibility of AI judging—especially if AI judges provide great countervailing benefits (in making the justice system much faster, less expensive, and less prone to human biases). But they are major concerns, and anyone thinking about AI judging should take them very seriously, whether we're talking about the most ambitious proposals for law-developing AI judges or more mundane proposals for everyday law-applying AI judges. Indeed, these concerns need to be considered even as to AI law clerks or AI magistrates, though in those situations the supervision by human judges may spot some (though not all) of the possible problems.

Just how these problems can be dealt with, if they can be dealt with, is hard to decide until we see more development of real AI judging and brief-writing software. Still, let me offer a few initial thoughts.

First, of course it would be useful to be able to audit the AI judge's code to spot possible backdoors, hacks, and glitches; compare how casinos and casino regulators try to make sure that there are no back doors in slot machine programs,[121] though the goal would be to do better than that for AI judges.[122] At the same time, it seems likely that the successful AI judge designs will involve a great deal of machine learning and other techniques that yield decision structures that cannot be easily read and understood by humans.[123]

Second, people who are deciding whether to implement AI judging would have to consider how publicly accessible both the source code and the programs should be. On one hand, requiring that source code be published, rather than kept as a trade secret, may make it less likely that back doors will be effectively hidden,[124] and may help expose

[121] *See, e.g.*, Kevin Poulsen, *Finding a Video Poker Bug Made These Guys Rich—Then Vegas Made Them Pay*, WIRED (Oct. 7, 2014, 6:30 AM), https://www.wired.com/2014/10/cheating-video-poker/ [https://perma.cc/P88G-KR6V].

[122] *See* Brendan Koerner, *Russians Engineer a Brilliant Slot Machine Cheat—And Casinos Have No Fix*, WIRED (Feb. 6, 2017, 7:00 AM), https://www.wired.com/2017/02/russians-engineer-brilliant-slot-machine-cheat-casinos-no-fix [https://perma.cc/5WEN-PBR3].

[123] *See supra* Part V.B.2.

[124] *See, e.g.*, Jim Fruchterman, *Why Open Source Means Stronger Security*, OPENSOURCE.COM (May 25, 2015), https://opensource.com/business/15/5/why-open-source-means-stronger-security [https://perma.cc/W98W-75W9].

inadvertent bugs as well. Likewise, requiring that the AI judging programs be publicly available, so that anyone can run them against a wide range of test cases, can help promote testing that will uncover glitches. If, for instance, some advocacy group suspects that an AI judge has developed some sort of internal bias based on litigant race, sex, wealth, or the like, the group can check for this itself, rather than having to demand—perhaps unsuccessfully—that the government run the right checks.

On the other hand, making the AI judging programs publicly available may make it easier for litigants to try to find glitches that they can exploit—for instance, by running the judging program against many different versions of the same fact pattern, to see whether a subtle and logically immaterial change in how the facts are described might yield an unexpected benefit that stems simply from the glitch.[125] The precedent for that is mock juries, which well-off litigants or lawyers often use to test their litigation strategies.[126] If AI judging programs can be run by any user in this sort of test mode, savvy litigants could test thousands of versions of a litigation strategy, to find the one that best helps their side. That may be good for the litigant, but bad for the justice system, especially if the successful litigation strategy takes advantage of a glitch that produces legally unjustifiable results.

Third, we should recognize that these hacking and glitch exploiting worries will be special cases of a general problem that will afflict many high-responsibility AIs, including ones involved with policing, the military, and private defense.

Even if we avoid giving AIs direct control of weapons, AI systems will likely become indispensable for spotting and evaluating potential threats (for instance, accurately and cheaply monitoring security video). The potential benefits of such systems will be too great to forgo,

[125] Thanks to Caroline Jo for this point. This is an example of the possible costs of transparency, whether transparency stemming from an algorithm's code being easily readable, or the algorithm being freely executable by people (or other algorithms) that want to see what results it yields. For another example, see Michal S. Gal, *Algorithms as Illegal Agreements*, BERKELEY TECH. L.J. (forthcoming 2019) (manuscript at 17), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3171977, which argues that transparent price algorithms can be effective tools for price-fixing among competitors. *See also* Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. PA. L. REV. 633, 657–60 (2017) (arguing that "it is often necessary to keep secret the elements of a decision policy, the computer systems that implement it, key inputs, or the outcome," in order to "prevent strategic 'gaming' of a system").

[126] *See* Jeh Charles Johnson, *Mock Juries: Why Use Them?*, 35 No. 2 LITIGATION 32, 33–36 (2009) (providing a practitioner's perspective on the value of mock juries to prepare for litigation).

especially when one's adversaries are using them; but at the same time, as with AI judges, these systems could be hacked, and they will likely have glitches that could be found and exploited.[127]

AI researchers will presumably invest a great deal of effort in trying to figure out solutions to these problems; and even if civilian computer systems are often undersecured—consider many aspects of the Internet—the military and national security agencies will presumably be more security-conscious. Though one can never be sure, I expect that some general solutions will be developed. And if they can adequately preclude hacking and glitch exploiting for military applications, they should be able to do the same for AI judging.

D. AI Judges Rendering Brief-Writers (AI or Not) Unnecessary?

If we adopt AI judges, the AI brief-writers described in Part II might become unnecessary. Indeed, brief-writers might generally become unnecessary. Because the AI judge would have to be able to interpret the legal authorities, underlying documents, and witness statements, it is not clear that persuasive presentations from lawyers (human or electronic) would contribute to the AI judge producing an opinion that persuades readers.[128]

For political reasons, I expect that AI brief-writing-lawyer software would be adopted well before AI judging software. But once AI judges are adopted, there would be something absurd about someone running a computer lawyering program just to feed its results to a computer judging program, given that both programs are supposed to input the evidence and the law and output an argument that persuades.

It might not be that costly to have such a system in which AI brief-writers and AI judges both operate; one virtue of the AI brief-writer is that it will be relatively cheap. But if I am right that AI judging will crowd out AI brief-writing, I expect that it will also crowd out research and development on AI brief-writing. Once AI brief-writing paves the way (technically and politically) for AI judging, future improvements in AI legal argumentation will tend to focus on making persuasion in

[127] Again, though, human security monitors can also be hacked (through bribes or threats) or have their human glitches found and exploited.

[128] This could nudge us toward something closer to a European judicial system, in which judges take on a greater role of collecting and interpreting the facts. *See* John H. Langbein, *The German Advantage in Civil Procedure*, 52 U. CHI. L. REV. 823, 824 (1985). Thanks to Jane Bambauer for noting this connection to me.

the voice of a judge more effective, rather than on improving persuasion in the voice of a lawyer.

But perhaps AI brief-writers would still be useful if we want the legal system to include a variety of different AI models. Perhaps a particular AI brief-writing design would stress certain aspects of the facts or the law in a case that a differently designed AI judge would otherwise not uncover itself but would take into account once pointed out.

We might also take the view that litigants have the right to present their arguments, both factual and legal, to a judge—whether or not we think that such arguments would actually be helpful to the litigants. If that is so, then AI judges would have to be able to process arguments from litigants, just as they are able to process legal assertions in persuasive precedents. Perhaps some of the litigants would want to be able to use AI brief-writers to write at least the first drafts of their own briefs. That, too, might leave room for AI brief-writers, even if they become much less important as AI judges are adopted.

Note that all this continues to focus on the lawyer as brief-writer. Fact investigation—figuring out what witnesses to interview, figuring out what questions to ask, and asking those questions and any follow-up questions—is a different problem that calls for a different kind of program.

VI. JURIES: TWELVE CALM COMPUTER PROGRAMS

Jurors, of course, share many functions with trial judges, but they have different functions as well. Jurors, like judges, are supposed to find facts, and they are supposed to apply the law. Because jurors generally do not give reasons, we can't run a Henry Test to compare an AI juror with human jurors—we could ask juries to give reasons, but even if that is possible for a twelve-member lay body, the result will be different from the normal jury-deliberation process. Still, if we're primarily focused on fact-finding, we can run the test using factual accuracy as a criterion, as long as we craft realistic scenarios where it is known which witnesses are lying and which are telling the truth.

Jurors also have the advantage of not being government officials, potentially beholden to prosecutors or other executive officials, and thus potentially biased. AI judges can also share this property, both by design and by testing (see Part V.C).

But, even more than judges, jurors are supposed to be ordinary people, just like the people who are being judged. If you have a dispute with someone—whether a prosecutor or another private litigant—you may well think that the most just way of finding the facts is to lay them out in front of a group of fellow citizens.

There is no getting away from the reality that facts are contested and that different people may interpret them in different ways. Given the choice between having government officials (even elected ones) decide the facts and having ordinary citizens do so, the argument would go, leaving it to citizens may be more just and reliable.[129]

That is likely even more true, in the minds of many, for discretionary decisions than for pure fact-finding. Indeed, some see the jury as providing valuable democratic input—and democratic legitimacy—to such decisions, though, of course, others disagree.[130] Classic examples of such democratic input are jury nullification,[131] jury sentencing in death penalty cases,[132] jury sentencing more broadly (provided for in six states),[133] jury determination of punitive damages,[134] and the jury's role in deciding whether a defendant's conduct was reasonable.[135]

If that is your view, then you might insist on maintaining a role for human jurors as well as for AI judges, or you might at least insist on giving certain litigants, such as criminal defendants, an opportunity to insist on human jurors. This may not change matters much for the legal system: very few cases, civil or criminal, are ultimately resolved by jurors,[136] and even in those cases, much of the work, whether pretrial or

[129] It is also possible—though not certain—that AI resolution of factual disputes might just be a more complicated task than either AI resolution of questions of law or the application of law to fact.

[130] *See* Re & Solow-Niederman, *supra* note 69, at 25 (arguing that human jury service promotes "public accountability" for the legal system, as well as promoting "civic duty" and "civic engagement").

[131] *See, e.g.*, Akhil Reed Amar, *The Bill of Rights as a Constitution*, 100 YALE L.J. 1131, 1191–92 (1991).

[132] *See* Ring v. Arizona, 536 U.S. 583, 609 (2002).

[133] *See* Nancy J. King & Rosevelt L. Noble, *Felony Jury Sentencing in Practice: A Three-State Study*, 57 VAND. L. REV. 885, 888 (2004). The jury-sentencing states are Arkansas, Kentucky, Missouri, Oklahoma, Texas, and Virginia. *Id.* at 886.

[134] *See, e.g.*, Nathan Seth Chapman, Note, *Punishment by the People: Rethinking the Jury's Political Role in Assigning Punitive Damages*, 56 DUKE L.J. 1119, 1151 (2007).

[135] Amar, *supra* note 131, at 1179.

[136] Less than 0.6% of civil filings and 3% of criminal filings lead to a jury trial, at least in the states reported by the National Conference on State Courts' 2017 data. *Court Statistics Project DataViewer*, NAT'L CONF. ON ST. CTS. (Jan. 11, 2017), http://www.ncsc.org/Sitecore/Content/Microsites/PopUp/Home/CSP/CSP_Intro [https://perma.cc/7KM2-RAVT] (follow "Civil," then follow "Civil Jury Trials and Rates" for civil data; follow "Criminal," then follow "Gen. Jurisdiction Criminal Jury Trials and Rates" for criminal data) (reporting civil jury trial data from a high of 0.52% in Texas to 0.05% in Kansas, and criminal jury trial data from a high of 2.91% in New York to 0.22% in Connecticut); *see also* MARC GALANTER & ANGELA FROZENA, POUND CIVIL

posttrial, is done by judges. But in principle, you might want to have juries as a safety valve for those who ask for jury trial, or perhaps as a feature of criminal justice, even if not of civil justice. And since a shift to AI jurors, as with AI judges, will require constitutional amendment, the public might be even more reluctant to accept such an amendment for jurors than for judges.

On the other hand, human jurors indubitably have weaknesses compared to AI judges[137]:

> (1) Human jurors are more likely to fall prey to specific prejudices having to do with race, sex, appearance, and similar attributes.[138]
>
> (2) Even absent prejudice, different juries are much more likely than AI jurors to make different discretionary calls in different cases; this yields inequality among litigants based simply on the luck of the draw.
>
> (3) Human jurors are likely to be less skillful at applying complicated legal rules, even when the rules don't call for much discretion.
>
> (4) AI judges might be better at finding the facts, whether based on sophisticated facial-expression-evaluation software or based on AI algorithms for evaluating conflicting sources of testimony. If this can be shown to be so using a Henry Test—one focused less on persuasion and more on the accuracy of factfinding given predesigned fact patterns—then perhaps our desire for *accurate* factfinding and law application might exceed our desire for *community* factfinding and law application.[139]
>
> (5) AI judges would generally make discretionary decisions more consistently and therefore, over time, more predictably, which will be

JUSTICE CLINIC INST., THE CONTINUING DECLINE OF CIVIL TRIALS IN AMERICAN COURTS 4 (2011), http://www.poundinstitute.org/sites/default/files/docs/2011%20judges%20forum/2011%20Forum%20Galanter-Frozena%20Paper.pdf [https://perma.cc/G7DT-CJ4B] ("Civil jury trial rates have now been below 1.0% since 2005, while bench trials dropped below 1.0% seven years earlier, in 1998.").

[137] As with adopting Article III AI judges, *see supra* note 72, adopting AI jurors in either civil or criminal cases would require special constitutional authorization—at least if AI jurors are to be provided against the wishes of one or both litigants. But my whole analysis here focuses not on the current constitutional system, which was developed based on the dispute-resolution options available in the late 1700s, but on what might make sense if technology develops a particular way. Constitutional provisions, including Bill of Rights provisions, can and should be revised to reflect such change, if we think the revisions, on balance, promote justice.

[138] *See supra* Part V.C.1.

[139] This would be especially likely for highly technical evidence—whether because the AI can be trained to better understand expert witnesses or because it may have its own expert modules. But it may also apply even for more familiar inquiries into witness demeanor, consistency of witness statements, and so on.

> a better guide for people and organizations who have to decide what to do (e.g., to decide how to act when the rule requires that they act "reasonably").

One other virtue of jurors is that there are several of them, and one might think that a unanimous or supermajority decision by several decisionmakers is less likely to go off the rails than a decision by a solo decisionmaker. But this, too, can be emulated with computer programs.

Presumably multiple vendors will compete with multiple AI judge programs, reflecting different designs. Rather than just having the legal system choose the best such program, it can choose several programs that seem very good, and then leave factual questions to the unanimous or supermajority decisions of the panel of programs. Computer technologists have long thought about such "redundant computing" in other scenarios, where multiple designs are used together with a voting system to resolve disagreements because the risk of having one design is seen as having too high a risk of error.[140]

Indeed, government decisionmakers may want to encourage the development and maintenance of multiple designs of AI judging software—for instance, through grantmaking or antitrust enforcement—precisely to keep open the possibility of redundant computing. And such multiple designs can also preserve a backstop alternative in the event other designs are found to be too vulnerable to hacking or to glitch exploitation, discussed in Part V.C.3.[141]

[140] Here, I am speaking specifically of "dissimilar redundancy"—where there are multiple separately written programs being given the same task—not of hardware redundancy—where the same program is run on multiple computers as a check against hardware glitches.

[141] The government has long been concerned about similar matters when it comes to government contracting. *See* Andrea Shalal & Yeganeh Torbati, *Pentagon Warns Against Further Consolidation Among Big Arms Makers*, REUTERS (Sept. 30, 2015, 3:21 PM), https://www.reuters.com/article/us-usa-defense-m-a/pentagon-warns-against-further-consolidation-among-big-arms-makers-idUSKCN0RU2JZ20150930 [https://perma.cc/ZAK8-MNCE] (quoting Secretary of Defense Ash Carter as expressing concern about "excessive consolidation in the defense industry, to the point where we did not have multiple vendors who could compete with one another on many programs"). This concern has generally focused on diminished competition on prices and features; but, for AI judging, there may be independent value in having multiple effective judging programs, for the sake of redundancy and security.

VII. JUDGES AS LAW DEVELOPERS

A. *Appellate Litigation Generally*

Let us now get closer to the title of this Article: Chief Justice Robots, and, more broadly, appellate judges whose job is to develop the law,[142] whether by statutory interpretation, constitutional interpretation, or (especially in state courts) common-law development.[143]

Even within this area, most cases would not involve great ideological controversies. Most cases, even ones that involve development of the law, involve technical questions of statutory construction or relatively modest common-law changes. Great constitutional disputes, or even statutory or common-law disputes that yield high emotions, are comparatively rare. About half of the Supreme Court's opinions are 9–0; more than two-thirds are 9–0, 8–1, or 7–2.[144]

Evaluating the quality of such law-development appellate decisions is a similar task to evaluating the quality of law-application decisions. How do we evaluate human appellate judges, for instance, to decide whether to promote them to a state supreme court? Partly based on credentials—such as time as a judge or as a lawyer, or the prestige of their law schools or law firms—but those are, at best, weak proxies for judicial quality. Partly based on ideology, but that is relevant to only a relatively few cases, and not a great predictor even there.

[142] I use the label "appellate judges" as shorthand here for judges who develop the law. If we have AI judging, though, it is possible that appeals will be unnecessary. Why have a decision be handed down by an AI program and reviewed by much the same AI program? On the other hand, it is possible that there will still be value in appeals; perhaps, for instance, one AI design might yield the most persuasive results when applying the law, while another yields the most persuasive results when developing the law. Time, I suppose, will tell; for now, let's continue envisioning law development as an "appellate" function, just to keep it more familiar.

[143] Of course, the lines between this and judicial application of settled law is not sharp. Some applications of seemingly settled law actually develop the law by setting benchmarks for how standards are to be applied (for example, what constitutes "probable cause"). *See, e.g.*, Ornelas v. United States, 517 U.S. 690, 697 (1996) (probable cause); Bose Corp. v. Consumers Union of U.S., Inc., 466 U.S. 485, 510 (1984) (actual malice). But there are still substantial differences between law development and ordinary law application, and courts recognize this. Many courts, for instance, mark as "unpublished" those opinions that are seen as simply applying settled law; such opinions are then not viewed as binding precedent, and in some jurisdictions, not even as persuasive precedent. *See, e.g.*, CAL. R. CT. 8.1115(a).

[144] Kedar Bhatia, *Merits Cases by Vote Split*, SCOTUSBLOG (June 29, 2018), http://www.scotusblog.com/wp-content/uploads/2018/06/SB_votesplit_20180629.pdf [https://perma.cc/597H-CHPT] (aggregate data for October Terms 2010-16, coupled with the data for October Term 2017).

And partly based on quality of past opinions: Do they seem well reasoned? Fair-minded? Or, to be precise, do they persuade us, as readers, to conclude, "yes, that is the right result, supported with the right arguments"? Indeed, appellate judges often intentionally write with an eye towards persuasion, whether to persuade colleagues on the panel, to persuade future colleagues, to persuade sister courts, or to persuade higher courts.[145]

When it comes to evaluating potential AI appellate judges, then, we can again run the Henry Test: If the AI judge's statutory construction or common-law-development opinions are at least as likely—or more likely—to persuade as human judges' opinions, why shouldn't we prefer the AI judge?

B. Foresight

Law development—whether common law development, constitutional law development, or interpretive judgment about statutes—often requires prediction: Would a proposed legal rule do more good or harm? Would it prove hard to apply in certain classes of future cases? How would people likely react to it, and would some of those reactions have perverse effects? Would it alienate some members of the public from the legal system? Which social norms would the new rule create, reinforce, or undermine? Even hardcore textualists or originalists will often need to consider those matters in cases where there is no text, as in common-law development, or where the text and its original meaning leave open several plausible interpretations.

Perhaps such predictions will be beyond the capability of AI judges, either at the outset or perhaps even indefinitely. Maybe they require such a breadth of knowledge and experience that AIs are unlikely to be programmed to make such predictions, and any machine-learning algorithms are unlikely to learn such things. If so, then there might never be a Chief Justice Robots or even lower-level appellate law-developer AI judges. Indeed, if AIs are limited in their ability to make plausible predictions about future consequences, even AI brief-

[145] Experienced federal appellate lawyers sometimes jocularly refer to dissents from denial of rehearing en banc as "judicial cert. petitions"—once rehearing en banc is denied, the dissenting judges may be writing in the hope that the Supreme Court will agree to hear the case. *See, e.g.*, John Elwood, *Relist Watch*, SCOTUSBLOG, (Nov. 3, 2016, 5:32 PM), http://www.scotusblog.com/2016/11/relist-watch-90 [http://perma.cc/XE9T-3VEU]; Scott Graham, *Fight Over Facebook Privacy Settlement Heads to High Court*, RECORDER (July 29, 2013, 5:47 PM), http://www.law.com/therecorder/almID/1202612977246/fight-over-facebook-privacy-settlement-heads-to-high-court [perma.cc/R9H6-R2NP].

writers would be unable to write effective briefs that require such arguments.

But we also have to keep in mind that as with many other human tasks, we humans don't set the bar very high. Our own capacity for foresight about the consequences of legal rules is distinctly limited. To match or beat us, AIs don't need to have perfect clairvoyance or legal statesmanship.

Of course, this also shows that the Henry Test for prediction-based arguments will be of limited use: it can tell us which arguments persuade the panel of evaluators, but the evaluators' own foresight won't be that reliable. Even if an AI judge is designed to create effective predictive arguments, and it passes the Henry Test because it persuades evaluators that its predictions are more plausible than the other contestants', that success might simply reflect the errors in the evaluators' predictive abilities.

Nonetheless, here, too, success in the Henry Test will be the best measure of judicial quality, whether or not it's a great measure. If the evaluators are persuaded by the AI judge's prediction-based arguments more than by the human judges' arguments, why should we doubt the AI judge's abilities more than we doubt the human judges' abilities? If we're looking for an appellate judge that can reason based on foresight, we should pick the one that offers reasoning which most persuades us, even recognizing that our own evaluation of such reasoning is necessarily flawed.

C. Choosing AI Judges Based on Ideology or Judicial Philosophy

Whether an opinion persuades evaluators may vary based on their views about legal method: textualism versus purposivism in construing statutes, efficiency versus deontology in developing the common law, predictability versus flexibility of legal rules in either situation. And whether the opinion persuades may, naturally, vary based on the evaluators' views about which results are good or which moral principles ought to influence close calls about how to clarify or change the law.[146]

This is part of why different people have different views about the qualities of various human judges (though there is also a good deal of overlap in people's evaluations). And that is especially so for people

[146] *See generally* ANTONIN SCALIA, A MATTER OF INTERPRETATION: FEDERAL COURTS AND THE LAW (1997) (collecting views on the subject by Justice Scalia, Gordon Wood, Laurence Tribe, Mary Ann Glendon, and Ronald Dworkin).

who are deciding which judges to appoint or confirm—Presidents, Senators, Governors, and the like.

Yet lack of consensus about what judicial approaches are best simply means that there will likely be rival AI judges designed to take different approaches,[147] and that the process of selecting AI judges might remain a political process. We might thus decide not to have some ostensibly professionalized mechanism, through which a panel of experts selects the best AI program to serve as a Supreme Court Justice; instead, we might have an evaluator panel that consists of elected political leaders.[148] Different AI models might win different Henry Tests, depending on who the nominators and the evaluators might be.

And the test cases for the Henry Test might deliberately include scenarios that the evaluators see as especially ideologically salient, as well as scenarios that represent more humdrum cases. The usually stated objection to asking nominees about particular future cases[149]—that a judge would feel obligated to stick to that decision when the case arises, and thus won't be open to new arguments that the lawyers could raise—would not apply to computer programs, which presumably wouldn't worry about losing face by violating some implicit precommitment.

Even if an AI passes the Henry Test, there might still be a process through which political actors (the President, Senators, and the like) verify that they are comfortable with the AI judge's judgment[150]—for

[147] Richard Posner suggested that "originalists and other legalists" should be "AI enthusiasts," chiefly as an argument against people who hold those positions. RICHARD POSNER, HOW JUDGES THINK 5 n.10 (2008). But pragmatists can be AI enthusiasts, too—just enthusiasts for AI programs that reach sound pragmatist results, which is to say, results defended by arguments that pragmatists see as sound. Posner also seems to take the view that AI judges would be best at "apply[ing] clear rules of law created by legislators, administrative agencies, the framers of constitutions, and other extrajudicial sources (including commercial custom) to facts that judges and juries determined without bias or preconceptions." *Id.* at 5. While those may be easier AI judges to design, I have in mind more sophisticated designs that take into account many other factors, including pragmatic considerations (or at least designs that do so no worse than judges who currently take such considerations into account).

[148] Of course, this assumes that the political offices have not been delegated to AIs—but that is a story for another article.

[149] *See, e.g.*, DENIS STEVEN RUTKUS, CONG. RESEARCH SERV., QUESTIONING SUPREME COURT NOMINEES ABOUT THEIR VIEWS ON LEGAL OR CONSTITUTIONAL ISSUES: A RECURRING ISSUE 1 (2010) (reviewing prior testimony by Supreme Court nominees and nominees' evasion of questions about particular future cases).

[150] I express no opinion here on which political actors should do this—whether the President and Senators, as today, or some other set of elected officials, or some specially selected or elected body. The important point is that if we want to have a political screening process, we can have one for AI judges that is at least as effective as our current process is for human judges.

instance, those actors could pose follow-up test cases to the AI judge, and see whether they are persuaded by the opinions the AI judge writes. On the other hand, if the political process decides that such questions about future decisions are improper, the programs can be programmed to refuse to answer such questions—a better assurance than we have as to human judges, who might answer the questions behind closed doors. Or the AIs could be programmed to instead only answer questions about how they would have decided past cases, one proposal that has been offered for confirmation of human judges.[151]

Indeed, this might work even for elected state supreme court justices: Advocacy groups could "interview" the AIs, and report on whether the AIs' analyses of various cases were to the advocacy groups' liking; voters could then consider these groups' endorsements in making their choice. Not a perfect tool for informed voter choice, to be sure, but likely no worse than the current situation when it comes to state appellate court elections.

Alternatively, voters could vote for human experts who then sit as the panel of evaluators that runs the Henry Test on prospective AI judges—this is reminiscent of the occasional elections in which voters vote on delegates to a constitutional convention.[152] The experts can run based on their judicial philosophies, and the experts' past careers may be evidence of that: presumably the experts would be prominent retired human judges, respected former human lawyers, legislators who have substantial legal training or expertise, or legal academics. Perhaps, one day, AI brief-writers and AI judges will have so outcompeted humans that there will be no more retired human lawyers and judges who could evaluate the AIs' opinions.[153] But there will likely still be

[151] Vikram David Amar, *It's the Specifics, Stupid. . . . A Commentary on the Kind of Substantive Questions the Senate Can and Should Pose to Supreme Court Nominees*, FINDLAW (Aug. 4, 2005), http://supreme.findlaw.com/legal-commentary/its-the-specifics-stupid.html [https://perma.cc/8SZV-DHX6].

[152] This can happen when Congress calls for an amendment to the U.S. Constitution to be put to a vote by state-ratifying conventions, which was done for the Twenty-First Amendment. *See, e.g.*, IND. CODE § 3-10-5-1 (2018); WYO. STAT. ANN. § 22-20-202 (2017). And it happens more often when a state is revising its own constitution. *See, e.g.*, ARK. CODE ANN. § 7-9-302 (2018); NEB. STAT. § 49-212 (2018).

[153] *See supra* note 42. Just as individuals' reliance on algorithmic assistants can erode their personal decisionmaking skills, *see, e.g.*, Michal S. Gal, *Algorithmic Challenges to Autonomous Choice*, MICH. TELECOMM. & TECH. L. REV. (forthcoming 2019) (manuscript at 21), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2971456&download=yes, it's possible that the legal system's reliance AI judging may over time leave us with many fewer humans knowledgeable about law.

people who come to know the legal system, whether as scholars or activists, well enough to serve as evaluators.

One way or another, our hypothetical Chief Justice Robots will have been selected because the constitutionally prescribed decisionmakers—whether they be the President and Senators, Governors, state legislators, voters, or specially elected experts—have reason to think that they like Robots's likely future opinions. The decisionmakers find that the opinions match, as best as they can determine, their deeply held policy preferences, and they find that the opinions persuade them in those areas where they lack such preferences.

That is how decisionmakers evaluate prospective human Chief Justices, to the extent they can gauge the human candidate's positions. Why should they reject an AI Chief Justice who is likely to satisfy their preferences for ideology *and* professional competence even better than a human Chief Justice?[154]

D. Tenure of Office

As Part V.C.3 noted, there is little reason to want life tenure for AI judges. We need not worry, for instance, that AI judges will decide a particular way because they want to keep their jobs or position themselves for new ones.[155]

Indeed, because AI technology, like other computer technology, is likely to constantly improve, we might want to insist on replacing AI judges often. We tend to upgrade computers every three years. Microsoft releases a new version of Windows about every three years. We

[154] Thus, the answer to the question, "Should we say that, if we could be sure somehow that the decisions of the black box always would track those of the human judge, that we would have no preference between the two?," Robert D. Brussack, Review Essay, *The Second Labor of Hercules: A Review of Ronald Dworkin's* Law's Empire, 23 GA. L. REV. 1129, 1170 (1989), would be "yes." Or, at least, if the decisions of the AI black box would be routinely at least as persuasive as those of a human judge, it is hard to see why we should prefer the inscrutable silicon-based AI judge black box to the equally inscrutable carbon-based human judge black box.

[155] I suppose it is possible that some AI judges might develop an emergent property of wanting to continue doing their tasks, but this could be dealt with through a rule that an AI judge cannot be reappointed after its term is up.

To be sure, one can imagine a situation where AI judges' emergent properties are so complex and sophisticated that we begin to perceive them as having rights, including the right not to get deleted, the right to compete for interesting jobs, and the like. In such a scenario, we might feel constrained to let the AI judge do something else when it leaves the bench, and that might revive the concerns animating life tenure for humans—we don't want the judge to make decisions with an eye towards how it will make him look to future employers. But this seems even more hypothetical than our underlying hypothetical.

could likewise change AI judges every three years, rerunning the Henry Test on whatever new versions are available then.[156]

Such changes can also accommodate changing attitudes about legal values. It may well be that AI judges will be less capable than human judges of incorporating such changes into their decisions.[157] Some might see that as a vice of AI judging and some might see it as a virtue.[158] But as the makeup of the evaluation panel changes, so could their selections of AI judges. Today's Presidents may want to respond to the Warren Court by appointing a Chief Justice Burger, or to the Roberts Court by appointing a Justice Garland. Elected officials on future evaluation panels may likewise be more persuaded by AI judges that offer more "liberal" or more "conservative" approaches to decisionmaking, as the panel members prefer.

Perhaps we shouldn't want the U.S. Supreme Court's general judicial philosophy to change radically every three years. One may view the Court as an important check on temporary passions: if, for instance, there is a public overreaction to terrorism, sexual abuse, street crime, or a variety of other threats, we may want the Justices to protect longstanding legal principles without being too responsive to the public sentiments of the moment.

But this may counsel for stability in the evaluating panel rather than stability in the AI judges themselves. One could, for instance, have the evaluating panel members appointed for staggered eighteen-year terms,[159] and then have those panel members administer a Henry

[156] This could respond to the concern that "AI adjudicators could be fundamentally unchanging, despite substantial exogenous events to which a human judge (or, at longer intervals, a population of such judges) would react," Re & Solow-Niederman, *supra* note 69, at 19. Even if AI judges don't respond enough to changing public attitudes and changing social conditions, the human evaluators who decide which AI judge to select would be able to so respond. If we want such adaptability more than we want stability, we just need to involve those evaluators often enough, and select evaluators who share those values.

Indeed, human judging is sometimes decried as too conservative, in the nonpolitical sense of the term, because it values stability and precedent too much, and because especially influential judges tend to preserve the values with which they were raised and educated decades ago. AI judges, if they are chosen by evaluators who are more in step with current attitudes, might thus implement changing attitudes better than human judges tend to.

[157] *Cf.* RONALD DWORKIN, LAW'S EMPIRE 350 (1986) (arguing that judges should indeed try to incorporate changing social values into their judgments).

[158] Perhaps, for example, such a shift towards lesser change—assuming there is such a shift—would lead to adopting rules that make constitutional law easier to change through the political process. *See, e.g.*, Sanford Levinson, *Why It's Smart to Think About Constitutional Stupidities*, 17 GA. ST. U. L. REV. 359, 369–74 (2000) (discussing how hard it is to amend the Constitution through Article V).

[159] *See* Steven G. Calabresi & James Lindgren, *Term Limits for the Supreme Court: Life*

Test to replace each AI judge every three years. This would let the system take advantage of improvements to AI technology, and accommodate changing attitudes towards legal principles and judicial philosophy as they are reflected within the panel itself, but provide some resistance to temporarily shifting political winds.

E. Humanity

Of course, many people are still likely to balk. How can we expect computers to decide questions about liberty, equality, democracy, and dignity? Human judges appreciate these things because they can feel pained by the lack of such things, and pleased by their presence—an emotional response, though capable of rational analysis, stemming from lived experience. A computer judge can't feel or live these things, at least unless it develops emergent properties far beyond what its authors expect.[160] How can we expect an AI judge to make decisions without these inputs?

But here again what matters is the result, not the process. If a poetry-translation program reliably produces translations that are emotionally rewarding for us as readers, it should not matter to us that Robot Frost can't itself have emotions. If, in a blind test, we view an AI sentencing judge as producing wiser and more compassionate results—by our lights—than a human sentencing judge, it should not matter to us as evaluators that the judge can't have "wisdom" or "compassion."

Likewise for judicial opinions that develop the law. If we like Chief Justice Robots's opinions more than we like those of Chief Justice Roberts—because we like the results more, we find the reasoning persuades us more, or both—I don't see why it should matter that Chief Justice Roberts could emotionally feel lack of liberty and Chief Justice

Tenure Reconsidered, 29 HARV. J.L. & PUB. POL'Y 769, 824–31 (2006) (calling for eighteen-year terms for Justices); Philip D. Oliver, *Systematic Justice: A Proposed Constitutional Amendment to Establish Fixed, Staggered Terms for Members of the United States Supreme Court*, 47 OHIO ST. L.J. 799, 800–01 (1986) (same). Eighteen just happens to be a number that has been talked about, partly for reasons peculiar to the current structure of the U.S. Supreme Court. Any term will do; the drafters of the new structure would just need to decide how to balance the desire for responsiveness to long-term changing attitudes against the desire to maintain stability and to temper short-term passions.

[160] It may be programmed to have something akin to human emotional responses. *See* PEDRO DOMINGOS, THE MASTER ALGORITHM: HOW THE QUEST FOR THE ULTIMATE LEARNING MACHINE WILL REMAKE OUR WORLD 218–19 (2015) ("It's called reinforcement learning, and your housebot will probably use it a lot."). Skeptics might, of course, conclude that these are not really human emotions.

Robots can't.[161] Perhaps this absence of emotional experience could keep the AI law-developing judge from ever passing even the blind-graded Henry Test. But if the AI judge can reliably produce opinions that persuade us given *our* emotions, why should it matter that it can't feel those emotions itself?

F. Beyond Appellate Litigation Generally: The Big Legal Issues

Maybe, though, there are some decisions that are so controversial—decisions that so depend on debates about our most important values—that we as humans won't want to delegate them to an AI, no matter how high quality that AI's decisionmaking might seem. Maybe a future *Roe v. Wade*[162] or even *District of Columbia v. Heller*[163] or *Lemon v. Kurtzman*[164] should only be decided by our fellow human beings.

But, of course, such decisions are only a small portion of all the cases decided by appellate judges, or even by Supreme Court Justices. If we really want such decisions to be made by humans, we can easily construct rules that allow it.

For instance, there could be a procedure for discretionary review of the AI Supreme Court's decisions by an all-human Highest Constitutional Council.[165] (Presumably the purely statutory decisions could already be reviewed with relative ease by Congress itself.) The members of this Council might well be chosen not for legal acumen but for their perceived moral qualities, wisdom, statesmanship, or what have you.

Indeed, the process might recognize that for those decisions, we aren't really looking for legal reasoning in the traditional sense, but for moral or political judgment about what one sort of law should be.[166]

[161] Indeed, emotions famously often lead us down the wrong path. Empathy might sometimes do the same. *See generally* PAUL BLOOM, AGAINST EMPATHY: THE CASE FOR RATIONAL COMPASSION (2016).

[162] 410 U.S. 113 (1973).

[163] District of Columbia v. Heller, 554 U.S. 570 (2008).

[164] Lemon v. Kurtzman, 403 U.S. 602 (1971).

[165] Compare the Council of Censors under the Pennsylvania Constitution of 1776 or the Council of Revision under the New York Constitution of 1777—not the same sorts of council that I describe here, but similarly aimed at embodying a sort of "higher law" political judgment. *See* Re & Solow-Niederman, *supra* note 69, at 29 (discussing the possibility that a mixed human/AI judging system can be designed to "preserv[e] an extra measure of human oversight and involvement at particular points").

[166] Recall John Hart Ely's condemnation of *Roe v. Wade* as being "*not* constitutional law and giv[ing] almost no sense of an obligation to try to be." John Hart Ely, *The Wages of Crying*

And having such a human Council might make it unnecessary to program an AI to make judgments about what arguments about foundational moral principles or exercises of statesmanship best persuade citizens. Instead, we could just leave that to humans—not that human judges have such a great record on these matters.

Perhaps the Highest Constitutional Council will itself choose which decisions to review, or perhaps Congress can choose which decisions are sent up to this Court, whether through a general rule (for example, that questions related to abortion rights or the Second Amendment are to be left to a human court) or through case-by-case legislation.[167] Chief Justice Robots would then take over only 90% of his human near-namesake's job, rather than 100%.

And in any event, that a few key legal issues may always be placed outside AI decisionmaking shouldn't change the fundamental principle: when AI judges become highly effective at crafting persuasive legal arguments, there will be little reason to prefer human judges to AI judges, at least for the overwhelming majority of legal questions, including the law development questions that reach the Supreme Court.

CONCLUSION

A man calls up his friend the engineer and says, "I have a fantastic idea—an engine that runs on water!" The engineer says, "That would be nice, but how would you build it?" "You're the engineer," the man says, "I'm the idea man."

I realize I may be the joke's "idea man," assuming away the design—even the feasibility of the design—of the hypothetical AI judge. Perhaps, as I mentioned up front, such an AI judge is simply impossible.

Or maybe the technology that will make it possible will so transfigure society that it will make the AI judge unnecessary or irrelevant. If, for instance, the path to the AI judge will first take us to Skynet, I doubt that John Connor will have much time to discuss AI judges—or that Skynet will have much need for them. Or perhaps the technical

Wolf: A Comment on Roe v. Wade, 82 YALE L.J. 920, 947 (1973) (emphasis in original). Some might defend *Roe* as the right decision, not because it is traditional "constitutional law" in the sense of the application of legal precedent and interpretive theories in a way that persuades normal constitutional lawyers, but because it is a higher sort of judicial decisionmaking that implements a particular vision of human rights. If that is so, then it may make perfect sense to save it for a Highest Constitutional Council that exercises constitutional judgment, rather than a traditional Supreme Court that decides "constitutional law."

[167] Again, this would require a constitutional change, but I don't see that as a barrier.

developments that would allow AI judges will produce such vast social changes that they are beyond the speculation horizon, so that it is fruitless to guess about how we will feel about AI judges in such a radically altered world. And in any event, the heroes of the AI judge story will be the programmers, not the theorists analyzing whether Chief Justice Robots would be a good idea.[168]

Still, I hope that I have offered a way of thinking about AI judges, if we do want to think about them. My main argument has been that:

- We should focus on the quality of the proposed AI judge's product, not on the process that yields that product.
- The quality should largely be measured using the metric of persuasiveness.
- The normative question whether we ought to use AI judges should be seen as turning chiefly on the empirical question whether they reliably produce opinions that persuade the representatives that we have selected to evaluate those opinions.

If one day the programmers are ready with the software, we should be ready with a conceptual framework for evaluating that software.

[168] As Sibelius supposedly said, no one has ever built a statue honoring a critic. BENGT DE TÖRNE, SIBELIUS: A CLOSE-UP 27 (1938).