

AI/ESQ.: IMPACTS OF ARTIFICIAL INTELLIGENCE IN LAWYER-CLIENT RELATIONSHIPS

CHRIS CHAMBERS GOODMAN*

Abstract

Whether we admit it or not, lawyers increasingly are working with machines in many aspects of their practice and representation, and it is important to understand how artificial intelligence can assist attorneys to better provide justice while recognizing the limitations, particularly on issues of fairness. This article examines current and future uses of technology to address how identity influences decisions about charges, defenses, credibility assessments, and communications in lawyer-client relationships. The article recommends that lawyers take affirmative steps to interact with AI technology developers to serve the interests of justice and fairness more fully.

Introduction

Discussions of artificial intelligence (“AI”) often start with an “aura of infallibility,” an expectation that it can do no wrong, much like DNA evidence had before the general public became more aware of its limitations. Empirical research and even random searches show the variety of errors (sometimes comical; other times, racist and sexist) that AI makes. Machine learning is a process that is influenced by the data fed into it, and learning algorithms need to be adjusted to minimize these errors. Data scientists explain that errors and bias are not the algorithm’s fault, in other words. That aura of infallibility has significant impacts as legal professionals expand their use of AI.

* Professor of Law, Pepperdine University School of Law; J.D. Stanford Law School; A.B. cum laude, Harvard College. I wish to thank my son, Alex Goodman, a Symbolic Systems major at Stanford University, for talking through many aspects of artificial intelligence with me as I prepared this article. His research assistance, along with that of my dedicated law student research assistant Caleb Miller, proved invaluable to this project. I also would like to thank reference librarian Don Buffaloe and Stacey Nelson for the assistance of the Faculty Support Office. I particularly thank the editors and all members of the *Oklahoma Law Review* for putting on an outstanding Symposium and for treating all of the authors with such respect and hospitality. To all the other panelists from the symposium, I express appreciation for their questions and comments to strengthen all of the articles in this Symposium edition. I dedicate this article to attorneys like me who are not “digital natives,” in the hopes of inspiring them to continue to give tech a chance.

AI has great potential to assist lawyers, as well as the opportunity to replace (some) lawyers. It can identify and minimize bias in client intake and initial consultations, it can assess the uniformity of criminal charging decisions made by prosecutors, and it can help to diversify law firm ranks, judicial ranks, and even juror pools. AI can also reduce the impacts of implicit bias by providing a mechanism for enhancing empathy, and by expanding the scope of information that lawyers rely upon, to provide a diverse array of circumstances as “normal” so that lawyers can make more inclusive judgments about clients, opposing parties, and witnesses.

Part I of this Article provides definitions of some key terms useful for this discussion, an overview of ways that AI is currently being used in the legal profession (including for bail and sentencing decisions), and some predictions of future applications, based on an analysis of the types of work lawyers do and how well-suited each task is to automation. Part II addresses the role identity plays in attorney-client relationships and highlights the challenges implicit racial, gender, and cultural biases raise, in the office as well as in the courtroom. These challenges appear in four main areas: (1) decisions about which clients to take and which charges or defenses to pursue, (2) credibility assessments for clients and witnesses, (3) communication methods, and (4) perceptions of justice and fairness. This Part also analyzes the ways in which AI could help reduce the impact of implicit and explicit bias in client representation and other areas of the justice system, and provides a critique of current processes. Part III explains how AI might exacerbate bias, and it concludes the Article with suggestions for further research and study, recognizing the crucial role that lawyers need to play in developing, training, and re-training the artificial intelligence that can assist all of us in providing greater access to justice.

I. Artificial Intelligence in the Legal Profession

A. What Is Artificial Intelligence?

Lawyers need a working definition of what constitutes artificial intelligence. Unfortunately, those definitions can be elusive—in part because of how quickly the technology is advancing, and in part because of disagreements over how to define it.¹ One proposed draft for a federal

1. Iria Giuffrida, Fredric Lederer & Nicolas Vermerys, *A Legal Perspective on the Trials and Tribulations of AI: How Artificial Intelligence, the Internet of Things, Smart Contracts, and Other Technologies Will Affect the Law*, 68 CASE W. RES. L. REV. 747, 752 (2018) (explaining that the dictionary defines AI as “a branch of computer science dealing with the simulation of intelligent behavior in computers,” or as “[t]he capability of a

definition of AI is “systems that think and act like humans or that are capable of unsupervised learning.”² Another definition is simply “technologies that use computers and software to create intelligent, human-like behavior.”³ A third recognizes that AI, cognitive computing, and machine learning are “generally interchangeable terms that all refer to how computers learn from data and adapt with experience to perform tasks.”⁴ In summary, “AI covers a gamut of technologies from simple software to sentient robots, and everything in between, and unavoidably includes both algorithms and data.”⁵

Of course, these definitions contain terms that also need to be defined. Algorithms are software programs that provide a “set of software rules that a computer follows and implements” by analyzing data and following the pre-determined instructions.⁶ “Machine learning” means “[t]he capability of algorithms and software to learn from data and adapt with experience.”⁷ Natural language processing, used in such commonplace technologies as Apple’s “Siri” and Amazon’s “Alexa,” is one application of machine learning, where the machine learns to process words rather than computer code.⁸

Returning to the legal realm, this Article will focus on three main aspects of AI currently in use: predictive artificial intelligence, analytic artificial

machine to imitate intelligent human behavior,” and concluding that they are “at best misleading and functionally useless”).

2. Tom Krazit, *Updated: Washington's Sen. Cantwell Prepping Bill Calling for AI Committee*, GEEKWIRE (July 10, 2017, 9:51 AM), <https://www.geekwire.com/2017/washingtons-sen-cantwell-reportedly-prepping-bill-calling-ai-committee/>.

3. David Kleiman, *Demystifying AI for Lawyers: Supervised Machine Learning*, ARTIFICIAL LAW. (Sept. 28, 2018), <https://www.artificiallawyer.com/2018/09/28/demystifying-ai-for-lawyers-supervised-machine-learning/>.

4. THOMSON REUTERS, *READY OR NOT: ARTIFICIAL INTELLIGENCE AND CORPORATE LEGAL DEPARTMENTS 4* (2017), https://static.legalsolutions.thomsonreuters.com/static/pdf/S045344_final.pdf.

5. Giuffrida, Lederer & Vermerys, *supra* note 1, at 756.

6. *Id.* at 753.

7. THOMSON REUTERS, *supra* note 4, at 4.

8. Sean Semmler & Zeeve Rose, *Artificial Intelligence: Application Today and Implications Tomorrow*, 16 DUKE L. & TECH. REV. 85, 87 (2017) (“What makes natural language processing unique from standard machine learning is how the program interprets commands. Rather than breaking a set of commands into a string of symbols or computer code, systems that use natural language processing are trained to interpret and understand questions presented in plain English, or any other language, by analyzing the words, sentence structure, syntax, and patterns of human communications.”).

intelligence, and machine learning.⁹ Predictive AI is a little more advanced and less commonly used by attorneys so far. One use of predictive AI involves inputting court decisions that allow the AI to predict the outcome of future legal cases.¹⁰ There have been some instances of high success rates in predicting how a court would actually decide. One experiment involved replicating judges' reasoning in the European Court of Human Rights.¹¹

Analytic AI evaluates statistics on the success rates for lawyers against particular judges, as well as opposing counsel¹² and can help lawyers to predict probabilities of success in particular courts or counties. AI can "forecast" arguments to be made by opposing counsel and evaluate the strengths of legal briefs and written arguments. These applications enhance the legal analytic work that attorneys have been performing with their brains thus far.

One familiar example of machine learning is the use of predictive coding to assist in reviewing large quantities of documents in litigation, which has been used for decades (back when this law professor was still a litigator).¹³ Machine learning can be stand alone or supervised.¹⁴ The term "supervised" also needs to be defined. With machine learning, "supervised" refers to when the person provides guidance to the machine in the form of outcomes, such as giving the machine a number of faces to evaluate, and then telling the machine which faces are human and which are not.¹⁵ If the lawyer carefully selects the data for training the algorithm, and guides the machine as it processes information and makes connections then we call that process supervised machine learning.¹⁶ The feedback from the human *as the machine*

9. Damian Taylor & Natalie Osafo, *Artificial Intelligence in the Courtroom*, LAW SOC'Y GAZETTE (Apr. 9, 2018), <https://www.lawgazette.co.uk/practice-points/artificial-intelligence-in-the-courtroom-/5065545.article>.

10. Matthew Hutson, *Artificial Intelligence Prevails at Predicting Supreme Court Decisions*, SCIENCE (May 2, 2017, 1:45 PM), <https://www.sciencemag.org/news/2017/05/artificial-intelligence-prevails-predicting-supreme-court-decisions>.

11. Taylor & Osafo, *supra* note 9 (stating that the success rate of this particular study was 79%).

12. *Id.*

13. See Charles Lew, *Artificial Intelligence and the Evolution of Law*, FORBES (July 17, 2018, 8:00 AM), <https://www.forbes.com/sites/forbeslacouncil/2018/07/17/artificial-intelligence-and-the-evolution-of-law/#118dc44036ee>.

14. Harry Surden, *Machine Learning and Law*, 89 WASH. L. REV. 87, 93 (2014).

15. *Id.*

16. *Id.*

learns provides some oversight, and that is what is necessary in the technology world to define the learning as “supervised.”¹⁷

“Unsupervised” includes situations where the person selects the input data for the machine to use but does not provide any guidance on the outcome. It is surprising to some lawyers particularly that providing data without any guidance on which outcomes the data should establish is still considered to be “unsupervised.” Unsupervised machine learning also includes situations where a human does not select the inputs; instead, the computer mines the internet or other data sets for information and almost figures out the optimal responses for itself.¹⁸ This unsupervised technology can be effective, such as with programs that “provide free legal advice on civil matters” by taking like facts of similar cases and compiling likely outcomes.¹⁹

However, it is easy to imagine that bad advice can result from an oversimplification of the facts. While unsupervised machine learning has its strengths in many areas (like obviating the requirement of feeding in a labeled dataset for training), it still creates the appearance of racist and sexist results when the computer output reflects the majority of the data, without making any judgments about the accuracy or truth of the data it analyzes.²⁰ For instance, unsupervised machine learning algorithms searched and interacted on the internet in an effort to learn human language and very quickly “learned” some significant use of profanity.²¹ As discussed more fully in Part III, ensuring the “accuracy” and “fairness” (which is still being defined in terms of AI), of the underlying data from which the machine learns is a significant challenge to using AI in the legal field.²²

17. But this is only true if the lawyer carefully picks the input data for training the algorithm; otherwise, supervised machine learning can be just as bad as unsupervised in terms of risk potential.

18. Rachel Wilka, Rachel Landy & Scott A. McKinney, *How Machines Learn: Where Do Companies Get Data for Machine Learning and What Licenses Do They Need*, 13 WASH. J.L. TECH. & ARTS 217, 223 (2018).

19. Gary E. Merchant, *Artificial Intelligence and the Future of Legal Practice*, SCITECH LAW., Fall 2017, at 20, 22, https://www.americanbar.org/content/dam/aba/publications/scitech_lawyer/2017/fall/STFA17_WEB.pdf.

20. See Matt O'Brien & Dake Kang, *Artificial Intelligence Plays Budding Role in Courtroom Bail Decisions*, CHRISTIAN SCI. MONITOR (Jan. 31, 2018), <https://www.csmonitor.com/USA/Justice/2018/0131/Artificial-intelligence-plays-budding-role-in-courtroom-bail-decisions>.

21. Jane Wakefield, *Microsoft Chatbot Is Taught to Swear on Twitter*, BBC (Mar. 24, 2016), <https://www.bbc.com/news/technology-35890188>.

22. Giuffrida, Lederer & Vermerys, *supra* note 1, at 758 (“[O]ne of the most difficult issues inherent in AI is how to assure that the data used by a computer is in fact accurate. Not only is information originating on the Internet, such as on social media, often inaccurate,

B. Current Uses of Artificial Intelligence in the Law

Lawyers are using AI in numerous ways. As a recent industry guide listed, the following are current uses of AI in the law: contract drafting and review, digital signatures, contract management, legal and matter management, contract due diligence, expertise automation, legal analytics, task management, title review, and lease abstracts.²³ In addition, there is now software that reviews legal briefs for strengths, weaknesses, patterns, and connections, and that can suggest additional cases as well as analyze the vulnerability of certain arguments.²⁴ Other areas in which lawyers are using AI include intake, document management, litigation budgeting, and evaluation of scientific expert testimony, as well as in bankruptcy, immigration, estate planning, taxes, securities, and food and drug cases. Criminal courts increasingly use computer algorithms in bail decisions. Risk assessment tools like COMPAS and the Public Safety Assessment (PSA) give judges a risk score for each defendant, and judges use those risk scores in determining whether to release the defendant on bail or to hold her or him until trial for the sake of public safety.²⁵

The most common uses still seem to be for legal research, document review, and drafting of standard documents. Lawyers are all familiar with Westlaw research and its ability to provide good analogies. Now, updated systems can do much more than help lawyers find case authorities. For instance, Ross Intelligence has a program that not only performs legal research, but also drafts research memos.²⁶ These programs can analyze briefs and cases and suggest missing cases from a list of authorities of an existing brief. There is even one company that has created a “bad law bot” to

but the Internet also contains intentionally false data often spread extensively by ‘bots’ and similar technologies that run automated tasks—such as spreading inflammatory content—at a higher rate than humanly possible.”).

23. Keith Mullen, *Artificial Intelligence: Shiny Object? Speeding Train?*, RPTE EREPORT (Fall 2018), https://www.americanbar.org/groups/real_property_trust_estate/publications/ereport/rpte-ereport-fall-2018/artificial-intelligence/ (citation omitted).

24. *Id.*; see also Edgar Alan Rayo, *AI in Law and Legal Practice – A Comprehensive View of 35 Current Applications*, EMERJ (Apr. 24, 2019), <https://emerj.com/ai-sector-overviews/ai-in-law-legal-practice-current-applications/>.

25. See, e.g., Alex Goodman, *Fairer Algorithms in Risk Assessment* (June 2018) (unpublished manuscript) (on file with author).

26. *Artificial Intelligence (AI) for the Practice of Law: An Introduction*, ROSS INTELLIGENCE (Aug. 8, 2018), <https://rossintelligence.com/ai-introduction-law/>; see also Susan Beck, *Inside ROSS: What Artificial Intelligence Means for Your Firm*, LAW.COM (Sept. 28, 2016), <http://www.law.com/sites/almstaff/2016/09/28/inside-ross-what-artificial-intelligence-means-for-your-firm/>.

ferret out when case or statutory law might be in question, even though it has not explicitly been overruled.²⁷ CARA is another program that provides summaries of the law and research memos.²⁸

Document review and management along with electronic-discovery (“E-discovery”) and predictive coding, searches databases of information for keywords that the lawyers have identified. This search is referred to as “technology-aided review” and also uses natural language techniques.²⁹ It helps to reduce the time attorneys spend on document review by culling out unimportant and irrelevant material from large volumes. In addition, the software can be trained to prioritize relevant documents.

Another increasingly common use of artificial intelligence is with judicial and agency actions that rely upon computer software to analyze vast quantities of data for purposes like predictive policing, bail setting (like the COMPAS program described above), and other matters.³⁰ A juvenile court judge used IBM’s Watson technology to analyze data for juvenile offenders and to create a several-page summary for each offender so that the judge could (perhaps) render better decisions in the approximately five to seven minutes per case that he had allotted.³¹ Another use of this technology generates draft answers to complaints that have been filed in court.³²

Contracts is another area in which AI is outpacing human lawyers. LawGeex (which drafts contracts), and Beagle (which reviews and organizes contracts, primarily for non-lawyers) are two providers in this area of law. “Smart contracts” are another emerging technology that use an algorithm that “autonomously executes some or all of the terms of the agreement” and relies upon data inputs rather than human supervision.³³ In one study, twenty attorneys sought to review five nondisclosure agreements alongside AI.³⁴ Not surprisingly, the AI was faster, but what was surprising to some is that the AI was also more accurate: “the AI finished the test with an average accuracy

27. *Meet Our Newest Team Member, Bad Law Bot!*, FASTCASE, <https://www.fastcase.com/blog/badlawbot/> (last visited May 9, 2019).

28. Allison Arden Besunder, *Not Your Parents’ Robot*, N.Y. ST. B.J., Mar./Apr. 2018, at 20, 90-APR NYSTBJ 20 (Westlaw).

29. Sergio David Becerra, *The Rise of Artificial Intelligence in the Legal Field: Where We Are and Where We Are Going*, 11 J. BUS. ENTREPRENEURSHIP & L. 27, 39–40 (2018).

30. Mariano-Florentino Cuéllar, *A Simpler World: On Pruning Risks and Harvesting Fruits in an Orchard of Whispering Algorithms*, 51 U.C. DAVIS L. REV. 27, 35 (2017).

31. Besunder, *supra* note 28, at 21.

32. *Id.*

33. Giuffrida, Lederer & Vermerys, *supra* note 1, at 759 (emphasis deleted).

34. Kyree Leary, *The Verdict Is In: AI Outperforms Human Lawyers Reviewing Legal Documents*, FUTURISM (Feb. 27, 2018), <https://futurism.com/ai-contracts-lawyers-lawgeex>.

rating of 94 percent, while the lawyers achieved an average of 85 percent. . . . On average, the lawyers took 92 minutes to finish reviewing the contracts. . . . [and] AI, on the other hand, only needed 26 seconds.”³⁵

If AI can complete contract review (which many lawyers consider a boring task) more quickly and better, judicious use of AI can lead to changes in the practice of law by helping attorneys be more effective and efficient in engaging in the types of legal work that AI is not so well-suited towards, such as litigation strategy. Computers are not likely to replicate strategy and planning anytime soon, as the lawyer still needs to conduct complex legal reasoning.³⁶

C. The Futures of AI and Legal Tasks

The futures of AI and the law depend on numerous factors, such as technological advancements, the willingness of lawyers to become more technologically savvy and to test out new products and approaches, the expectations of clients, courts, and other actors in the legal market, and the extent of regulation (or not) and accuracy (or not) of the specific AI applications in use. At this point there is not enough data on how lawyers actually divide their tasks and time. Despite the perennial complaint about wasting time filling out timesheets to allocate one’s time every day in six-minute increments, full and accurate descriptions on timesheets are important for tracking billable hours in order to determine when and where AI can provide the most efficient assistance. Given clients’ increasing attention to billable rates, particularly for junior lawyers who are still basically “in training,” greatly increasing efficiencies in legal work is a business imperative for law firms. As the New York Bar Association cautions: “the billable hour as a measurement of value for a lawyer’s work has been long overdue for a disruption,” because a “computer never tires and will ‘brute force’ its way through massive amounts of data, without the need for an expensive dinner and a car service home.”³⁷

One group of researchers took on this task to figure out how much time lawyers spend on various categories of work. They used the resources of a consulting firm that analyzes lawyer invoices, and they evaluated thirteen

35. *Id.*

36. V.R. Feroze & Lorien Pratt, *How AI Is Disrupting the Law*, DIGITALIST MAG. (Apr. 3, 2018), <https://www.digitalistmag.com/digital-economy/2018/04/03/ai-is-disrupting-law-06030693>.

37. Besunder, *supra* note 28, at 23 (“If AI can take the robot out of the lawyer and make the practice more about the strategic and intellectual analysis, then we should not necessarily ‘fear the (AI) reaper’”).

categories based on these invoices to create a distribution of hours, billed by task, and then allocated percentages of overall lawyer billing time.³⁸ The thirteen categories were document management, case administration and management, document review, due diligence, document drafting, legal writing, legal research, legal analysis and strategy, fact investigation, advising clients, negotiation, other communications and interactions, and court appearances and preparation.³⁹ The study focused on Tier 1 (over 1000 lawyers) to Tier 5 (under twenty-five lawyers) firms and did not include solo practitioners or contract attorneys (which constitute a significant number of the overall lawyers working for clients).⁴⁰

The largest amount of lawyer time in all categories was on legal analysis and strategy at 27% and 28.5%.⁴¹ The second largest category (depending on firm size) was with either legal writing or court appearances and preparation.⁴² Fact investigation and advising clients, as well as other interactions and communications, were mainly in the third position.⁴³

Document management and legal research were at less than 1% of the overall time lawyers billed, which may be surprising to some.⁴⁴ One explanation for this low impact could be that so many of these tasks are performed by paralegals or clerical staff or are being automated already.⁴⁵ Computer coding has been used in the legal field for decades to automate searches and to categorize documents for production and use at trial. Similarly, computers are more well-suited to doing certain types of legal research quickly, as one Miami litigator learned when he spent about ten hours searching to discover a case whose facts mirrored what he was working on, while his firm's AI software found the same case almost instantly.⁴⁶

Based on their description of various types of artificial intelligence and the limits of automating legal work, the authors of that study provide a prediction as to the future automation impacts for certain tasks. For instance, not surprisingly, they predict document management will have a light

38. Dana Remus & Frank Levy, *Can Robots Be Lawyers? Computers, Lawyers, and the Practice of Law*, 30 *GEO. J. LEGAL ETHICS* 501, 506–07 (2017).

39. *Id.* at 508.

40. *Id.*

41. *Id.*

42. *Id.*

43. *Id.*

44. *Id.*

45. *Id.* at 513–14.

46. Holly Urban, *Artificial Intelligence: A Litigator's New Best Friend?*, *LAW TECH. TODAY* (Oct. 15, 2018), <https://www.lawtechnologytoday.org/2018/10/artificial-intelligence-a-litigators-new-best-friend/>.

employment impact in part because it is such a low percentage of the time attorneys bill for.⁴⁷ They predict that using AI in legal writing will have a weak employment impact on human lawyers because so much of legal writing is not structured in a way that can be adequately automated,⁴⁸ despite the contract drafting software discussed above. Similarly, advising clients, fact investigation, negotiation, and court appearances and preparation also have weak employment impacts, as they too are difficult to do by machine.⁴⁹ The areas in which they expect a moderate impact are case administration and management,⁵⁰ due diligence,⁵¹ document drafting,⁵² legal research,⁵³ as well as legal analysis and strategy.⁵⁴

In terms of litigation strategy, some artificial intelligence systems analyzing vast quantities of data can be trained to identify trends and data points that do not fit in, much the way that Columbo, from the 1970s, 80s, and 90s television detective series, focused on that which did not make sense—facts and anomalies that (almost) always turned out to be the key to enable him to “crack the case.”⁵⁵ This type of intelligence will be useful in assisting attorneys and investigators with developing litigation strategies. In the courtroom, judges can use it for collating the relevant cases quickly and easily, to help them save time in rendering their judicial decisions. In addition, courtrooms could efficiently use speech recognition software, such as Dragon for dictation (once expanded to allow for multiple users) to provide real-time transcripts at a low cost and much faster speed than human court reporters.⁵⁶

Artificial intelligence as a means of conflict resolution or adjudicating legal cases is another area of penetration in the legal environment. A computer science researcher decided to try to reproduce a human judge’s ruling in a dispute over a Dodgers home run ball. In that case, the fan who

47. Remus & Levy, *supra* note 38, at 508 (noting that it is less than 1% for those at the Tier 1 to Tier 5 firms).

48. *Id.* at 519.

49. *Id.* at 525–29.

50. *Id.* at 514–15.

51. *Id.* at 517–18.

52. *Id.* at 518–19.

53. *Id.* at 520–23.

54. *Id.* at 524–25; *see also* W. Bradley Wendel, *The Promise and Limitations of Artificial Intelligence in the Practice of Law*, 72 OKLA. L. REV. 21, 23–24 (2019) (describing what tasks computers are good at executing).

55. Taylor & Osafo, *supra* note 9.

56. Jonathan Lounsberry, *Using Technology Inside & Outside the Courtroom*, FAM. ADVOC., Spring 2015, at 8, 12.

caught it was hit by a mob, which caused the ball to fall out of his hand, so that another fan retrieved it.⁵⁷ The computer determined that while it would be reasonable for the first fan, Popov, to be deemed the owner of the ball, the other fan, Hayashi, would be an acceptable owner as well.⁵⁸ The human court reached the same conclusion. As a brief *Kansas Bar Journal* article noted, “the most equitable solution was the Solomonic order to have the ball sold with the proceeds of the sale divided evenly.”⁵⁹ This application of machine learning for dispute resolution is expanding through websites like OneDayDecisions.com, which charges about \$50 to decide a small claims-type court case.⁶⁰

Another area where artificial intelligence can be useful is in providing legal assistance to those who cannot afford it, and perhaps better representation than would otherwise be available. For instance, as Professor Henderson suggests, “AGI criminal defense lawyers could bring human-level—or even superhuman—competence to every minute (and even every microsecond) of every representation.”⁶¹ AI also can provide language interpretation so that litigants can properly exercise their rights even if they do not speak English.⁶² As California Supreme Court Justice Cuéllar notes:

If society enhances the artificially intelligent tools available for addressing challenges of such enormous legal consequence, we will gain new opportunities to close the considerable gap

57. Larry N. Zimmerman, *Artificial Intelligence in the Judiciary*, KAN. B.J., July/Aug. 2016, at 20, 20.

58. *Id.*

59. *Id.* The author, Zimmerman, notes that “[t]he computer’s wishy-washy result feels like a cop-out but it mirrored the California Supreme Court’s reasoning. The human court held that both Popov and Hayashi had legal rights to the ball and neither could be lawfully deprived of it.” *Id.*

60. Zimmerman surmises that the artificial agents will become assistants or support to the adjudicators. *Id.*

61. Stephen E. Henderson, *Should Robots Prosecute and Defend?*, 72 OKLA. L. REV. 1, 16 (2019). Professor Henderson explains that the “transaction costs of trying to find competent counsel if one has some money to pay,” as well as the concerns of some defense lawyers about whether they can in good conscience defend a particular person, suggest to him that “while there does not seem to be inherent reason to prefer human defense counsel, there does seem inherent reason to retain human prosecutors,” notwithstanding instrumental benefits. *Id.* at 16, 18.

62. Cuéllar, *supra* note 30, at 36 (“Some kinds of software applications leveraging artificial intelligence – whether expert systems, generic algorithms, or neural nets – can help address all the problems we currently face living up to these ideals. We do not have enough interpreters or lawyers to help poor people. Identifying the right molecules to regulate is difficult.”).

between legal aspirations and reality that currently bedevils aspirations for justice. Convolutional neural networks and certain kinds of expert systems with natural language user interfaces grafted on can help with interpreting, facilitate legal advice, and enhance the capacity of agencies to discern what they should regulate.⁶³

Access to justice can be greatly enhanced, but we must remain mindful of whether legal instruments are valid and enforceable, especially as more people turn to self-help AI. Disregarding the lawyer entirely may not be enough to protect people's interests and serve their legal goals, as Professor Poppe discusses in the area of trusts and estates.⁶⁴ Justice Cuéllar cautions:

[E]xisting law may encourage automation without some careful weighing of aggregate risks or consequences. We are in fact faced with an exceedingly blurry line between computer-assisted human choice and human-ratified computer choice. We can begin to see the complexity of this question by looking to older cases examining liability for both excessive reliance and insufficient reliance on computing systems.⁶⁵

It is one thing for artificial intelligence to assist attorneys in making better, or fairer, or more efficient judgments, but it is a different situation where the human is simply ratifying what the computer has chosen to do.

The machines may be very good at telling the human lawyers what to do and when to do it, but not necessarily how to do it (strategy) or why (ethics). Regulating online legal providers and figuring out the line between legal assistance and authorized practice of law is another significant issue that lawyers need to resolve as technology grows more prevalent.⁶⁶ There also may be perverse incentives in the AI for private actors to overcomply (such as with TurboTax to avoid consumers being audited), and for governmental actors to overcomply (to avoid challenges in administrative hearings, for

63. *Id.* (citation omitted).

64. Emily S. Taylor Poppe, *The Future is ~~Bright~~ Complicated: AI, Apps & Access to Justice*, 72 OKLA. L. REV. 183, 199 (2019) (warning that “[w]hile greater interaction may increase the quality of the final product, it may also increase the likelihood that courts will find these programs to be instances of UPL”).

65. Cuéllar, *supra* note 30, at 39.

66. Susan Saab Fortney, *Online Legal Document Providers and the Public Interest: Using a Certification Approach to Balance Access to Justice and Public Protection*, 72 OKLA. L. REV. 91, 120-21 (2019) (concluding that “[a] certification system promotes both transparency and consumer choice”).

instance), as discussed more fully in Professor Morse's paper in this symposium issue.⁶⁷

In summary, artificial intelligence provides a number of benefits for lawyers and law firms. Specifically, AI can help attorneys be more efficient by permitting them to focus on their creative analysis rather than the tedious and often frustrating or stressful aspects of their work.⁶⁸ Because AI cannot replicate the human interaction aspect of lawyers as counselors and advisors in the myriad of circumstances that may present themselves, it frees up the lawyer's time to spend on having more effective client relationships.⁶⁹ Thus lawyers still will be needed to do the work for which AI is not, and likely will not be, well-suited.⁷⁰

II. The Roles of Identity (Race and Gender) in Lawyer-Client Relationships, and How AI Can Help Lawyers Reduce Bias

This Part will describe ways in which race, gender, and socio-economic status, as well as other aspects of identity, impact attorney-client relationships. First, these identities play a role in determinations about which clients the lawyers will represent, which defendants the prosecutors will charge, and how and whether defense attorneys will represent them. Second, these characteristics can shape credibility assessments and strategic choices that lawyers make throughout the relationship. Third, these demographic characteristics affect communications, both substantively and procedurally. AI can provide assistance to reduce these manifestations of lawyer bias.

The fourth role that identity plays is in the perception of justice and fairness of the entire legal system. AI can assist in hiring and promoting diverse lawyers and judges and in selecting more diverse juries, all of which

67. Susan C. Morse, *When Robots Make Legal Mistakes*, 72 OKLA. L. REV. 211, 213-14 (2019). Professor Morse explains how government robots might produce decisions that undercomply with the law, thus benefiting recipients of public aid because decisions that overcomply with the law and therefore deny such benefits, would be challenged, likely in administrative proceedings. On the other hand, she notes that for market robots "such as TurboTax, the situation is reversed," because overcompliance means the consumer is not audited, whereas undercompliance could result in government challenges to the accuracy of the tax returns. *Id.*

68. Avaneesh Marwaha, *Seven Benefits of Artificial Intelligence for Law Firms*, LAW TECH. TODAY (July 31, 2017), <https://www.lawtechnologytoday.org/2017/07/seven-benefits-artificial-intelligence-law-firms/>.

69. *Id.*

70. Wendel, *supra* note 54, at 24-25 (noting seven areas in which computers are unlikely to be able to replace lawyers in the foreseeable future, including fact investigations, negotiation over settlement terms and work in "new or rapidly changing areas of law").

on their face can help promote the perception of justice and fairness. AI can also be re-trained, more easily perhaps than humans can, to address any biased outcomes once discovered.

A. Using AI to Choose Which Clients and Which Cases

In determining which cases to accept in private practice and which to prosecute or defend in the criminal realm, lawyers make judgments about the person standing before them, whether that person is a potential client, a person who may be charged with a crime, or a person whom the attorney must defend (as in the case of public defenders without any choice in whom they represent).⁷¹ As lawyers obtain more experience making these decisions, they use that experience to learn and to adapt their decision-making process.⁷²

This learning can include racial and gender factors, which may be based on generalizations and stereotypes as well as on actual experience.⁷³ Generalizations form because they are sometimes, usually, or often true, at least in our experiences.⁷⁴ Stereotypes are positive and negative characterizations based on cultural cues and context, as well as the history of dominance and oppression in the given society.⁷⁵ Human actors do not necessarily rely upon racial factors intentionally because implicit and

71. *People v. Barnett*, 954 P.2d 384, 457 (Cal. 1998) (arguing that guilt could be based on a “gut feeling”).

72. Allison C. Shields, *How to Avoid Bad Clients Before They Enter Your Practice*, LEGALEASE (Oct. 3, 2014), https://www.legaleaseconsulting.com/legal_ease_blog/2014/10/how-to-avoid-bad-clients-before-they-enter-your-practice.html (explaining how to choose clients that are best for a firm, by creating a checklist of reasons not to accept a potential client, the first being, “Something about the client makes you uncomfortable”). This common line of reasoning is how implicit bias prevents those of different racial, gender, cultural, or economic backgrounds from receiving proper legal aid.

73. See generally Chris Chambers Goodman, *The Color of Our Character: Confronting the Racial Character of Rule 404(b) Evidence*, 25 LAW & INEQ. 1 (2007) [hereinafter Goodman, *The Color of Our Character*] (describing the dangers of courtrooms using racial generalizations to support character assumptions, either without, or alongside suspect Federal Rule of Evidence 404(b) admissions); see also Chris Chambers Goodman, *Shadowing the Bar: Attorneys' Own Implicit Bias*, 28 BERKELEY LA RAZA L.J. 18 (2018) [hereinafter Goodman, *Shadowing the Bar*].

74. See Goodman, *The Color of Our Character*, *supra* note 73, at 14 (noting a “particular individual may have had some past experience with [race] that confirms the ‘truth’ of this stereotype for him”).

75. John F. Dovidio et al., *Prejudice, Stereotyping and Discrimination: Theoretical and Empirical Overview*, in THE SAGE HANDBOOK OF PREJUDICE, STEREOTYPING AND DISCRIMINATION 3, 7 (John F. Dovidio et al. eds., 2010) (defining a stereotyping as “the typical picture that comes to mind when thinking about a particular social group”).

unconscious bias can operate as well.⁷⁶ Subtle cues provide social context, and this is what enables humans to learn racism.⁷⁷

AI can streamline the client selection process through automated intake. It can create screening questions to assess quickly and more accurately the potential conflicts (as already happens in larger firms), but at an even deeper level, such as by identifying all the potential witnesses and making connections (much like Facebook and LinkedIn provide connections to whomever you might know) that a paralegal may overlook.

In another context, those who represent clients based on a contingency fee pricing structure improve at selecting which clients and cases are more likely to provide victories and larger settlements or awards upon trial, or they go out of business. AI can provide a more detailed and potentially more accurate evaluation of the likelihood of winning at trial. Although it is not a foolproof assessment, it can support a decision as to whether to accept the representation. Some lawyers may be using AI to mass market to potential clients, and these uses of AI will increase, as Professor Bernstein explains in her article in this Symposium issue. When taken to extremes, however, such uses could promote gender and other biases in these marketing campaigns.⁷⁸

In addition, individual clients who earn more than 125% of the poverty level may need artificial intelligence because otherwise they may not have access to attorneys.⁷⁹ In fact, “[f]or most individuals, the choice is not between a technology and a lawyer. It is the choice between relying on legal

76. Matthew T. Nowachek, *Why Robots Can't Become Racist and Why Humans Can*, PHAENEX, Spring/Summer 2014, at 57, 79. As Nowachek explains, “[R]acism need not necessarily function as a chosen disposition, though more often than not it is coexistent with an unconscious skillful conformity to a racist norm that is passed through embodied mediums as seemingly banal as language, gestures, eye contact, or body positioning.” *Id.*

77. *Id.*

78. Anita Bernstein, *Minding the Gaps in Lawyers' Rules of Professional Conduct*, 72 OKLA. L. REV. 123, 131 (2019).

79. *Frequently Asked Questions*, AM. BAR ASS'N (July 16, 2018), https://www.americanbar.org/groups/legal_services/flh-home/flh-faq/ (noting that the American Bar Association has provided free legal aid and pro bono programs for people with incomes “less than 125 percent of the federal poverty level”). Many states and large counties have similar state-run programs. *See, e.g., How Do I Qualify?*, LEGAL AID FOUND. L.A., <https://lafla.org/help/qualify/> (last visited May 10, 2019); *see also Federal Poverty Guidelines – 2019*, MASSLEGAL SERVS (Jan. 11, 2019), <https://www.masslegalservices.org/content/federal-poverty-guidelines-2019> (showing that 125% of the poverty level is a mere \$15,613 annually for a one-person household). Absent a contingency fee agreement, a lawyer could not expect to be paid for his services.

technologies or nothing at all.”⁸⁰ They cannot afford moderately priced attorneys and are not eligible for free services.⁸¹

Similarly, prosecutors learn and adapt to better decide which cases they should prosecute, which they should end with a plea bargain, and which they should allow a jury to decide. Much of this learning process in humans is based on their interaction in the real world, and that real world includes the race, gender, and other cues that provide context for processing information.⁸² AI can access past cases, plea bargains, and conviction rates to recommend which charges, and at what level, the prosecutor should file.

With this information, AI can help the lawyers themselves to identify bias and potential bias, rather than simply masking names, deleting photos and gender references, or removing other demographic data. One way AI helps lawyers identify bias mathematically is to de-correlate variables like zip code and last name from socioeconomic status.⁸³ AI can provide an “objective” decision or analysis,⁸⁴ which the attorney can then compare with his or her own analysis or decision process, such as selecting which clients to represent and which witnesses to put on the stand to testify. The San Francisco District Attorney recently announced that his department will implement this program for charging decisions, de-correlating information explicitly about race and ethnicity, as well as information that is indirectly or implicitly racial,

80. Tanina Rostain, *Robots Versus Lawyers: A User-Centered Approach*, 30 GEO. J. LEGAL ETHICS 559, 568–69 (2017) (explaining that even with the eligibility cap for free legal services set at 125% of the poverty level, there are only enough poverty lawyers to help half of those who are qualified). Therefore, those above the poverty level must pay for services or go without.

81. *Id.*

82. Nowachek, *supra* note 76, at 77–78 (“[T]hrough specific physical habits that a human being develops within a cultural context he or she is able to tap into the common-sense knowledge of that context in a manner that bypasses any difficulties with determining the relevancy of racial factors.”).

83. Flavio P. Calmon et al., *Optimized Pre-Processing for Discrimination Prevention* (n.d.), <https://papers.nips.cc/paper/6988-optimized-pre-processing-for-discrimination-prevention.pdf> (paper presented at the 31st Conference on Neural Information Processing Systems (NIPS 2017)).

84. Tonya Riley, *Get Ready, This Year Your Next Job Interview May Be with an A.I. Robot*, CNBC (Mar. 13, 2018, 10:28 AM), <https://www.cnbc.com/2018/03/13/ai-job-recruiting-tools-offered-by-hirevue-myra-other-start-ups.html> (quoting Lindsey Zuloaga, director of data science at HireVue) (“We can measure it, unlike the human mind, where we can’t see what they’re thinking or if they’re systematically biased . . .”).

such as surnames, hair and eye colors, and neighborhoods.⁸⁵ Once the assistant district attorney makes a decision about how and whether to charge the case, the program will disclose the redacted information for the attorney to consider.⁸⁶ If the attorney wishes to change her charging decision after reviewing the information correlated with race and ethnicity, she must document the reasons for the change.⁸⁷ Of course, the attorney may simply rationalize a non-racial reason for the change in charging decisions, as research studies suggest people often do after the fact to conform with their predetermined conclusions.⁸⁸ Nevertheless, this AI program is a good first step in identifying bias.

Another method is to have AI make the decision in parallel with the human as noted above, but to use counterfactuals in its analysis. For instance, the AI can be instructed as follows: “all the facts of the case are the same, but let’s pretend your client is a white man instead of a black woman,” and see if the machine-predicted outcome now differs. Some frequently proposed solutions for reducing implicit bias include raising awareness to motivate people to self-monitor.⁸⁹ However, some people, despite an explicit warning not to use race, will then do just that. Making race central to the analysis can have positive and negative impacts, and it is unclear which would be more significant.⁹⁰

85. *All Things Considered: San Francisco DA Looks to AI to Remove Potential Prosecution Bias* (National Public Radio broadcast June 15, 2019), <https://www.npr.org/2019/06/15/733081706/san-francisco-da-looks-to-ai-to-remove-potential-prosecution-bias>.

86. *Id.*

87. James Queally, *San Francisco D.A. Unveils Program Aimed at Removing Implicit Bias from Prosecutions*, L.A. TIMES (June 12, 2019), <https://www.latimes.com/local/lanow/la-me-san-francisco-da-prosecutions-implicit-bias-software-20190612-story.html> (“If a prosecutor’s decision changes between the two phases, they will be expected to document what led to the change ‘in order to refine the tool and to take steps to further remove the potential for implicit bias to enter our charging decisions,’ according to a statement issued by the district attorney’s office.”).

88. Jerry Kang et al., *Implicit Bias in the Courtroom*, 59 UCLA L. REV. 1124, 1156-59 (2012) (describing an experiment where employers decide which of two candidates they prefer first, and then are asked to justify whether experience or education is the more important factor in the decision). The authors conclude that “[t]his research suggests, however that implicit motivations might influence behavior and that we then rationalize those decisions after the fact. Hence, some employment decisions might be motivated by implicit bias but rationalized post hoc based on nonbiased criteria.” *Id.* at 1159.

89. Natalie Salmanowitz, *Unconventional Methods for a Traditional Setting: The Use of Virtual Reality to Reduce Implicit Racial Bias in the Courtroom*, 15 U. N.H. L. REV. 117, 129 (2017).

90. *Id.*

B. Credibility Assessments and AI

Lawyers often rely upon “intuition” developed by experience in evaluating which potential witnesses, and even clients, are being truthful and forthright.⁹¹ Intuition is also referred to as a “gut” feeling, and is described by many as “one of those things I just know.”⁹² As the science of implicit bias has developed, we have learned that much of what we purport to “just know” is based on our preconceived notions, biases, and prejudices, whether we are aware of them or not.⁹³ For this reason, it is important to de-construct decision-making processes that rely upon intuition, to determine whether those decisions are influenced by biases that may negatively impact the lawyer’s assessment of one witness’s credibility or positively impact an assessment of another witness’s credibility, without a principled reason for actually making a distinction in the credibility analysis for each witness.

For example, one way to assess credibility is for the lawyer to ask, “Does this story ring true? Does it make sense based on what I know about how the world operates?”⁹⁴ Questions like this may seem to be an appropriate mechanism for an initial up or down decision about whether a witness is being truthful. And in the real world, these questions may work to help us figure out which of two children broke the expensive vase—the one who says, “I didn’t break it. I heard the crash when I was playing outside,” or the other one who says, “I didn’t break it. I saw a monkey hanging from the ceiling when I walked into the room. He must have done it and then ran away.” Unless we live near a zoo with faulty animal security, the second story does not make sense. However, if we live in a jungle, or even near one of the parks filled with monkeys in the city of Delhi, India, the second story also makes perfect sense. We determine which story makes sense based on our knowledge of the world and our experiences.⁹⁵ When lawyers have

91. Andrew T. Barry, *Selecting Jurors*, LITIG., Fall 1997, at 8, 11 (“[E]xperienced jury lawyers all admit to relying on feelings or instinct in gauging the predispositions of individual jurors.”); *id.* at 63 (“[I]ntuition[] and insights should be focussed less on picking favorable jurors, and more on dismissing unfavorable ones.”).

92. See *People v. Barnett*, 954 P.2d 384, 457, (Cal. 1998) (prosecutor arguing that it was his “gut feeling” that the defendant was guilty during closing argument).

93. Anthony G. Greenwald & Linda Hamilton Krieger, *Implicit Bias: Scientific Foundations*, 94 CALIF. L. REV. 945, 951 (2006).

94. L. TIMOTHY PERRIN, CAROL A. CHASE & H. MITCHELL CALDWELL, *THE ART AND SCIENCE OF TRIAL ADVOCACY* (2003).

95. Matthew D. Lieberman, *Free Will: Weighing Truth and Experience*, PSYCH. TODAY (Mar. 22, 2012), <https://www.psychologytoday.com/us/blog/social-brain-social-mind/201203/free-will-weighing-truth-and-experience> (“Experience outweighs truth nearly every time.”).

different experiences and live in almost different worlds from the witnesses, based on race, gender, socio-economic status, national origin or other characteristics, they may too quickly discount stories that do not make sense in their worldview, without taking the time to assess the witness's credibility from the perspective of that witness's worldview.

The real culprit here is the lawyer's inability to hear the story from another perspective, from another point of view.⁹⁶ Of course, in litigation as well as negotiation, lawyers need to convince others of their clients' stories; thus, perhaps the lawyer is right to disregard a story that is not going to make sense to the opposing party, mediator, judge, or jury. In an individual case, this assessment may be true. But the systematic lack of confidence in the credibility of those who come from different places, different realities, and different experiences, leads to the fourth issue—undermining the goals of justice and fairness for clients and witnesses of all types and from all places.⁹⁷

How can AI help with credibility assessments? First, AI can track eye movements, facial cues, changes in voice and tone, and heart rates, much like a more sophisticated version of a lie detector test. Second, AI can help lawyers to better assess the credibility of facts and stories by expanding the realm of what is possible (such as asking Siri about the possibility of loose monkeys in the example above), or providing actual facts (such as weather forecasts and almanac data) to prove that an alibi is credible (for instance, that a storm did knock out the road, and that the extra driving distance under those traffic conditions means that the client could not have been present at the time and place of the crime). Furthermore, AI can identify the trends in data that do not fit, and find the thread that unravels the entire scheme, a la Detective Columbo.

C. What Messages Are Lawyers Conveying and Receiving?

Third, communications in the office and in the courtroom can be influenced by implicit racial, gender, and other biases.⁹⁸ Cultural biases may be held by judges, jurors, litigants, and their lawyers.⁹⁹ Because all of these

96. See generally Sonia N. Lawrence, *Cultural (in)Sensitivity: The Dangers of a Simplistic Approach to Culture in the Courtroom*, 13 CAN. J. WOMEN & L. 107 (2001)

97. See *id.*

98. Goodman, *Shadowing the Bar*, *supra* note 73, at 30 (describing how differences between cultures cause people to take different meaning from the same set of facts, and to react differently from the same verbal cues). This misunderstanding of one another often goes unnoticed as we assume the other party to the conversation understood what we said as we understood it. *Id.*

99. Kang et al., *supra* note 88, at 1135–51 (describing implicit racial and cultural bias present in criminal prosecutions from the initial police interaction, to the judgment and jury).

legal actors bring their backgrounds with them into the courtroom, there is always a danger that these biases may play out in the many levels of communication in a legal case.¹⁰⁰ For instance, lawyers can “prime” jurors negatively to act on biases and prejudices using coded rather than explicit language.¹⁰¹ They can also inoculate jurors from those pre-existing biases and prejudices¹⁰² by confronting the potential biases explicitly.

Both inoculation and priming matter, given that studies have shown that jurors tend to reach a decision before the actual formal deliberation process begins.¹⁰³ During deliberations, the jurors’ knowledge of cultural and other stereotypes (including racial and gender stereotypes) has an impact on fellow jurors and on their conversations.¹⁰⁴ The communications that judges provide can be similarly helpful or harmful through jury instructions, their stern or

100. Mark B. Bear, *Injustice at the Hands of Judges and Justices*, PSYCH. TODAY (Apr. 15, 2017), <https://www.psychologytoday.com/us/blog/empathy-and-relationships/201704/injustice-the-hands-judges-and-justices> (“[Judges’] decisions are based upon their personal biases, beliefs, assumptions and values, which are formed as a result of our personal backgrounds and life experiences [and] [w]e all have personal biases . . .”).

101. Kathryn M. Stanchi, *The Power of Priming in Legal Advocacy: Using the Science of First Impressions to Persuade the Reader*, 89 OR. L. REV. 305, 308 (2010) (defining priming) (“[T]he ‘prime’ or stimulus (the words or information about [a subject]) ‘excites’ an area of the brain that contains information about a particular category . . .”). For example, if one were to say words like “Juvenile,” “Thug,” “Aggressive,” to a jury before seeing a black teenage defendant, she will have likely primed their implicit bias against that demographic with words frequently used describing criminal behavior against them. *See also* Jesse Marczyk, *The Adaptive Significance of Priming*, PSYCH. TODAY (Jan. 22, 2017), <https://www.psychologytoday.com/us/blog/pop-psych/201701/the-adaptive-significance-priming> (defining priming as “an instance where exposure to one stimulus influences the reaction to a subsequent one”).

102. *Reptile Defense Tactics: Priming in Voir Dire*, EXCELAS (Aug. 25, 2016), <https://www.excelas1.com/perspectives/blog/post/2016/08/25/Reptile-Defense-Tactics-Priming-in-Voir-Dire.aspx>. For example,

[a] plaintiff attorney uses the question “Who here feels that physicians should always put safety as their top priority?” to prime jurors to the concept of “safety.” The defense can attempt to re-prime jurors by instead asking, “Who here feels that a physician’s real priority needs to be treating every patient as a unique individual?” In this case, the defense can strip away the plaintiff attorney’s “safety” priming, and instead prime jurors to focus on the concept of treating a unique patient, rather than strict adherence to general safety rules.

Id.

103. Karenna F. Malavanti et al., *Subtle Contextual Influences on Racial Bias in the Courtroom*, JURY EXPERT (May 29, 2012), <http://www.thejuryexpert.com/2012/05/subtle-contextual-influences-on-racial-bias-in-the-courtroom/>.

104. *Id.*

quizzical facial expressions, and even their manner of ruling on objections.¹⁰⁵ So much of our legal arguments rely upon subtle as well as overt communications, and cultural biases can have a negative impact on serving the interests of justice.¹⁰⁶

Some would say that having a lawyer is like a gatekeeper, because a lawyer can provide access to justice, and not having a lawyer can impede or result in an outright denial of access to justice in some situations.¹⁰⁷ But the mere presence of a qualified lawyer is not the end of the analysis. What that lawyer says and does, as well as how she interacts with the client, opposing counsel, and the court, are all important aspects of the attorney-client relationship. When the lawyers have trouble walking in the shoes of their clients and identifying with someone who is “other,” the crucial bond of mutual trust may be lacking, further impeding successful outcomes for the client and for the justice system as a whole.¹⁰⁸

One option for using AI to assist with communications in the courtroom noted by researcher Natalie Salmanowitz is to screen judicial actors, such as prospective jurors and even judges, with the Implicit Association Test prior to trial.¹⁰⁹ Being made aware of their potential biases might help lawyers, jurors, and judges to act more deliberately to compensate for those biases. There could be some substantial opposition to this approach based on the critiques of the IAT (such as those about the adequacy of evidence of causation as opposed to simply correlation between high IAT scores and actual biased behaviors).¹¹⁰ In addition, giving the scores to other parties and their counsel could constitute a breach of privacy, or at best another socially undesirable

105. *Judges' Nonverbal Behavior in Jury Trials: A Threat to Judicial Impartiality*, 61 VA. L. REV. 1266, 1281 (1975) (writing that the judge’s non-verbal expressions “are an uncontrolled reflection of the judge’s own feelings about the case”). Further, “[d]uring the testimony the attitude of the judge is very important. His movements and gestures, even his posture, affect the jury and they react accordingly.” *Id.* at 1268.

106. Masua Sagiv, *Cultural Bias in Judicial Decision Making*, 35 B.C. J. L. & SOC. JUST. 229, 236 (2015) (recommending that cultural experts be used to testify at trial to help to mitigate judge’s biases).

107. See Herbert M. Kritzer, *Contingency Fee Lawyers as Gatekeepers in the Civil Justice System*, JUDICATURE, July-August 1997, at 22, 29.

108. *Id.*

109. Salmanowitz, *supra* note 89, at 123–28.

110. Goodman, *Shadowing the Bar*, *supra* note 73, at 26 (citing Maurice Wexler, *The Survival of the Intentionality Doctrine in Employment Law: To Be or Not to Be* (Nov. 2016) (unpublished manuscript)) (“Even if the IAT tests accurately predict the existence of pathways that evidence implicit bias, they do not demonstrate that people act consistently with the implicit biases that the test measures.”).

burden of mandated jury service.¹¹¹ Another option is to weaken stereotypical associations, which can be accomplished by providing counter-stereotypical examples, such as lifting up former President Barack Obama to counter the “violent” and “uneducated” stereotypes about African American men. AI can scour the internet to provide counter-examples quickly and thoroughly, but only if the right questions are asked (black world leaders, not black presidents, for instance), and the right training of the algorithm is performed as it learns. Because the examples need to be realistic, and preferably well known, in order to have the greatest impact, this approach has its limitations. Research is mixed on this proposition about weakening stereotypes because while some recent researchers were able to demonstrate reduction in stereotypes based on counter-stereotypical examples, other researchers report not being able to replicate those findings.¹¹²

The main criticisms of these proposals for reducing bias in lawyers are (1) that the tools will apply to very small numbers of people, and (2) that with some people, the tools may be counterproductive, thus exacerbating implicit racial biases.¹¹³ Salmanowitz has a bold proposal to make this solution more efficacious—using virtual reality or other video games with jurors and judges. She concludes that “virtual reality exercises have the potential to reduce implicit racial biases more effectively than measures proposed in existing literature,”¹¹⁴ and are therefore worthy of “serious consideration.”¹¹⁵

Although virtual reality is not the same as artificial intelligence, her proposal provides a useful illustration. Salmanowitz explains that virtual reality tools could be used to minimize implicit bias in judges if incorporated into the courtroom,¹¹⁶ although the theory has not been tested. Digital games could highlight counter-stereotypical situations and thus minimize or reduce the salience of race in the players. Judges or jurors who use these games may demonstrate somewhat reduced implicit biases,¹¹⁷ at least temporarily.

Another aspect of virtual reality involves immersive virtual environments including “body ownership illusions, in which individuals temporarily feel as though another person’s body part is in fact their own. Unsurprisingly, these

111. Salmanowitz, *supra* note 89, at 135.

112. *Id.* at 135-36 nn.135-45.

113. *Id.* at 136.

114. *Id.* at 160.

115. *Id.*

116. *Id.* at 146-47. The article explains the background that prior studies have shown factors like proximity to lunch time determine whether or not someone is going to be granted or denied parole, because of hunger and or fatigue, and recognizes that there are many extralegal factors that operate in the courtroom implicitly and explicitly. *Id.* at 118-19.

117. *Id.* at 140-41.

illusions are particularly effective in reducing self-other distinctions.”¹¹⁸ Some studies based on the image of an avatar have found that people with lighter skin who received an avatar with a darker skin tone demonstrated reduced implicit bias levels on the IAT after the fact.¹¹⁹ The authors caution that because these games and technology environments have not been developed for the explicit purpose of providing enhanced fairness and reduced bias, additional tests would be necessary to determine how effective the technology may be.¹²⁰ Still, an increase in empathy is a likely result.

The next Part considers the difficulties with the fourth aspect, perceptions of justice and fairness.

D. Justice and Fairness: Perceptions, Realities, and Predictions

Another way that artificial intelligence can reduce bias of lawyers is in the hiring process. It can detect patterns, which may demonstrate the existence of bias, and can be programmed to disregard demographic information like race and gender in making initial screenings of resumes, candidates, and applicants.¹²¹ For instance, in the hiring arena, there is a company called HireVue¹²² that analyzes assessment tools for potential employers and uses the expertise of corporate industrial psychologists, as well as AI, to help reduce bias. To the extent that companies use video conferencing for screening interviews, however, racial, ethnic, and appearance biases may not be mitigated at all. Another hiring company, Pymetrics, has developed a method for removing bias from their algorithms and provides that service to the public.¹²³ Others recommend using artificial intelligence to identify bias in order to “nudge us into being better people instead of doing the work for us.”¹²⁴

Another commonly suggested solution is to diversify the pool of lawyers and judges, as well as jurors, and the AI mechanisms used in hiring, selection, and employee assessment may help with lawyers and judges. Using AI to send out juror notices and consciously expand the pool of potential jurors with algorithms correlated to enhance diversity could make an impact

118. *Id.* at 141.

119. *Id.*

120. *Id.* at 143–45.

121. Chad Getchell, *Reducing Unconscious Bias with AI*, PEOPLES-COUT (Nov. 26, 2018), <https://www.peoplescout.com/reducing-unconscious-bias-with-ai/>.

122. HIRE VUE, <https://www.hirevue.com/> (last visited May 10, 2019).

123. PYMETRICS, <https://www.pymetrics.com/employers/> (last visited May 10, 2019).

124. Erika Morphy, *Can Artificial Intelligence Weed Out Unconscious Bias?*, CMSWIRE (Feb. 13, 2018), <https://www.cmswire.com/digital-workplace/can-artificial-intelligence-weed-out-unconscious-bias/>.

as well, though that effect can be mitigated if the lawyers exercise their challenges on those jurors during *voir dire*. Diversifying pools can have some impact, as studies have shown that white jurors in more diverse environments may, and in some instances, do make even greater efforts to avoid biased behavior.¹²⁵ The efforts to diversify jury pools could be significant, depending upon who is selected from the pool and how lawyers use their peremptory challenges and challenges for cause, but efforts to diversify judges would not make as much of a difference because in most cases only one judge hears the case.¹²⁶

One additional way that machines can help reduce the impact of lawyers' biases is that the machine can be re-taught, and in certain circumstances, learned biases can be unlearned or corrected. We have been trying to do that with people through the Implicit Association Tests as well as Elimination of Bias Continuing Legal Education courses for lawyers, but we still have miles to go. In contrast, correcting machine learning can be more certain and more successful.

In sum, people are human. Machines are not. Machines can be better at making cost-benefit analyses (provided they have the right data), they are not partial to themselves, and they have the ability to consider far more options much faster, and with greater ease, than a human can.¹²⁷ Artificial intelligence provides a mechanism for correcting some of the unfairness that humans engage in, both intentionally and unintentionally. By expanding examples, enhancing empathy, and highlighting the potentials for bias, AI can assist attorneys greatly in reducing the manifestations and effects of bias in law practice. While AI could help mitigate perceived biases in justice and fairness, it may actually replicate and thereby perpetuate injustice and unfairness if humans do not closely monitor and re-train it. Part III addresses and responds to this concern.

III. Overcoming Biases in Artificial Intelligence

A. In What Ways Might AI Maintain, or Even Increase, Bias?

It is crucial to understand that artificial intelligence requires training in order to learn how to identify patterns and do its job, and thus it has limits

125. Salmanowitz, *supra* note 89, at 132.

126. *Id.* at 137.

127. Michael Anderson & Susan Leigh Anderson, *Machine Ethics: Creating an Ethical Intelligent Agent*, AI MAG., Winter 2007, at 15, 18 (predicting future machine capability to self-regulate ethics).

based on the data inputs it receives and the people who help to build it.¹²⁸ The data accessed provided opportunities for algorithms to embed sexism and gender stereotypes in their results, such as when Amazon learned that its new recruiting system scanned resumes for notations like “Women’s Club” and determined that such notations were factors against a positive hiring decision.¹²⁹

The feedback loop that results from making decisions based on biased data can perpetuate bias, and embedded biases exacerbate those problems and project them into the future, potentially magnifying the harm, all because of the way the machine learns.¹³⁰ As Justice Cuéllar notes, it is only because we can reverse engineer the situation that we can understand the bias. The danger of not knowing how the machines reach their conclusions could lead to misappropriations of justice.¹³¹ The lack of transparency in machine learning algorithms is one of these dangers because people may not really know how the computer reaches its decision.¹³²

128. Tom Gummer, *Legal AI Beyond the Hype: A Duty to Combat Bias*, KENNEDYS L. (Sept. 26, 2018), <https://www.kennedyslaw.com/thought-leadership/article/legal-ai-beyond-the-hype-a-duty-to-combat-bias> (noting examples from Google News, such as “when asked to complete the statement ‘Man is to computer programmer as woman is to X’, replied, ‘homemaker’”). The article further discusses how AI can be used to help predict legal outcomes but cautions about monitoring the data set inputs, stating,

What if, for example, the training data consisted of 1,000 claims, 500 brought by women and 500 brought by men, but every one of the claims brought by a female was settled for under £20,000 and every one of the claims brought by a male was settled for over £20,000. Would that lead to the creation of a fair predictive solution? No.

Id.

129. Nicole Lewis, *Will AI Remove Hiring Bias?*, SHRM (Nov. 12, 2018), <https://www.shrm.org/resourcesandtools/hr-topics/talent-acquisition/pages/will-ai-remove-hiring-bias-hr-technology.aspx>.

130. Ben Dickson, *Artificial Intelligence Has a Bias Problem, and It’s Our Fault*, PC MAG. (June 14, 2018, 2:00 PM), <https://www.pcmag.com/article/361661/artificial-intelligence-has-a-bias-problem-and-its-our-fau>.

131. Cuéllar, *supra* note 30, at 33–34 (“The intricate pattern recognition made possible by these techniques comes at an analytical price, though, as in some cases it is far from clear, even to the designers of the systems, precisely how they have arrived at their conclusions.”).

132. Wendell Wallach, *Rise of the Automats*, 5 SAVANNAH L. REV. 1, 7 (2018) (“Among the more immediate concerns is the transparency of learning algorithms. Learning algorithms largely require massive data input from which specific outfits are determined. But no one fully understands what happens between that input and output. The algorithms can’t tell you, and the computer scientist can’t go back and tell you what the learning system has done. There is no transparency.”).

Can AI be or become biased? A group of researchers used machine learning with human language to demonstrate how a machine might learn language biases. They used the Implicit Association Test, developed a word embedding association test, and found that “machine learning absorbs stereotyped biases as easily as any other” type of learning.¹³³ They found that names associated with European Americans were more easily associated with positive characteristics than names associated with African Americans. In addition, they replicated the finding that female names and descriptions are more associated with family than career, and with the arts rather than the sciences. The researchers caution against using these artificial intelligence technologies to perpetuate cultural stereotypes that can result in prejudiced outcomes.

Nowachek argues that robots cannot *become* racist, because they have an inadequate relationship to the social world and to practice, but he also admits that they can *be* racist.¹³⁴ He explains that in order for there to be racism, there has to be some account of race, but because race is not biological and rather a social concept, it “operates as a type of value judgment concerning a person or a group of people” that really only makes sense “within a network of social relationships and practices in which values are socially constructed and assigned.”¹³⁵ As racism must operate within this social context, “it more often than not is expressed merely in terms of ‘what everyone simply knows.’”¹³⁶

He describes a play on words such as the use of the word “pen,” which would be difficult for a computer to process without the context of other

133. Aylin Caliskan, Joanna J. Bryson & Arbind Narayanan, *Semantics Derived Automatically from Language Corpora Contain Human-Like Biases*, 356 SCIENCE 183, 183 (2017).

134. See Nowachek, *supra* note 76, at 79. The author summarizes this reasoning as follows:

[R]obots cannot become racist insofar as their ontology does not allow for an adequate relation to the social world which is necessary for learning racism, where racism is understood in terms of a social practice. This is revealed most clearly in the failure of robots to manage common-sense knowledge in its tacit and social forms—a problem that has come to be known as the common-sense knowledge problem.

Id. at 58-59 (citing Hubert L. Dreyfus, *Why Heideggerian AI Failed and How Fixing It Would Require Making It More Heideggerian*, 171 ARTIFICIAL INTELLIGENCE 1137, 1138 (2007)).

135. *Id.* at 64.

136. *Id.* at 65 (quoting Errol Lawrence, *Just Plain Common Sense: The ‘Roots’ of Racism*, in CTR. FOR CONTEMPORARY CULTURAL STUDIES, *THE EMPIRE STRIKES BACK: RACE AND RACISM IN 70S BRITAIN* 47 (1982)).

words (is it a writing instrument or an enclosure for an animal?). Similarly, when a woman tenses up upon encountering a shabbily dressed man on a dark street late at night, the computer does not understand the context of sexual assault as a rationale for why that woman might have cause for concern.¹³⁷

Nor would the computer “understand” that she relaxed her tension when she realized the man was white, rather than black. As robots do not “just simply know,” Nowachek therefore concludes that “robots have no meaningful way to receive the racial cues necessary to become racist.”¹³⁸ This example begs the question that if machines could be trained to receive and process racial cues, would they also be able to determine which racial cues are relevant and/or appropriate and which are not?

Another challenge with artificial intelligence at this stage in its development is that it has no concept of the historical context that informs the data. For instance, the computer can do a historical analysis of all the presidential candidates for all the actual presidents of the United States, and it would conclude that one must be a white male in order to be president of the United States based on that data. The computer does not have the capacity to understand that during a certain part of our country’s history, nonwhite men were not permitted to vote and therefore were unlikely to be eligible to run for office or to be elected, nor that women did not obtain a vote for some significant period of time and therefore were not as likely to run for high office.

On the other hand, the forty-fourth President would be deemed an outlier, and the computer would disregard that data if it were evaluating the race, gender, or ethnicity of those who had the best chance of becoming president. “Blindness to bias is a fundamental flaw in this technology,” and fairness is not embedded.¹³⁹ In fact, “unlike humans, whose brains tend to notice and react strongly to ‘outliers,’ machine-learning algorithms tend to discount or ignore them.”¹⁴⁰ Hence the importance of having human beings be part of the decision-making feedback and re-evaluation and recalibration process. As one programmer notes, “we have to teach our algorithms which are good associations and which are bad the same way we teach our kids.”¹⁴¹ IBM is

137. *See id.* at 66-67.

138. *Id.* at 71.

139. Jonathan Vanian, *Unmasking A.I.’s Bias Problem*, FORTUNE (June 25, 2018), <http://fortune.com/longform/ai-bias-problem/>.

140. *Id.*

141. Dana Bass & Ellen Huet, *Researchers Combat Gender and Racial Bias in Artificial Intelligence*, BLOOMBERG (Dec. 4, 2017), <https://www.bloomberg.com/news/articles/2017->

making use of humans to develop methods to reduce bias in data sets and to enhance equity.¹⁴² As Justice Cuéllar notes, “[T]he Internet was designed to be adaptable and relatively resilient rather than secure.”¹⁴³

Another place where bias operates is that the machine may perceive data involving minorities to be “outliers,” because the minority groups represent such a low percentage of the overall data. Outlier data is or can be excluded, and then it is not factored into the machine learning process at all.¹⁴⁴ For instance, some studies find facial recognition software has a 2% accuracy rate for black, African, and African-Caribbean faces, and more than double that, but still only 5% accuracy for white men.¹⁴⁵

B. Proposed Interventions for Reducing Bias in AI

Providing additional data or more good data is one way to minimize biased feedback decision loops.¹⁴⁶ If the data does not include many examples of certain types of people (e.g., diverse candidates), then the computer program may not do a good job evaluating those diverse candidates, thus increasing the reliance upon humans (whom we already know are subject to biases as well).¹⁴⁷ Updating data sets could be a time-consuming process, but it is a preferred alternative to letting the algorithm learn from biased data.

A person could make decisions about which associations are appropriate and which are not, such as consciously adding photos of nonwhite males to lists of CEOs and other candidates for various positions,¹⁴⁸ and thus “train” the algorithm to be more inclusive of other races and genders, despite statistics to the contrary in the data set. Comparing the “objective” decisions

12-04/researchers-combat-gender-and-racial-bias-in-artificial-intelligence (quoting TOLGA BOLUKBASI ET AL., *MAN IS TO COMPUTER PROGRAMMER AS WOMAN IS TO HOMEMAKER? DEBIASING WORD EMBEDDINGS* (2016), <https://papers.nips.cc/paper/6228-man-is-to-computer-programmer-as-woman-is-to-homemaker-debiasing-word-embeddings.pdf>).

142. See *AI and Bias*, IBM, <https://www.research.ibm.com/5-in-5/ai-and-bias/> (last visited May 10, 2019).

143. Cuéllar, *supra* note 30, at 37. See generally Becerra, *supra* note 29.

144. Noel Sharkey, *The Impact of Gender and Race Bias in AI*, HUMANITARIAN L. & POL’Y BLOG (Aug. 28, 2018), <https://blogs.icrc.org/law-and-policy/2018/08/28/impact-gender-race-bias-ai/> (discussing the bias and automatic face recognition software noting that it is most accurate for white males and least accurate for darker female faces).

145. *Id.*

146. See Jason Bloomberg, *Bias Is AI’s Achilles Heel. Here’s How to Fix It*, FORBES (Aug. 13, 2018, 9:40 AM), <https://www.forbes.com/sites/jasonbloomberg/2018/08/13/bias-is-ais-achilles-heel-heres-how-to-fix-it/#28d5c4d26e68>.

147. See *id.*

148. See Dickson, *supra* note 130.

of the AI with the “subjective” decisions of the human could help the machine to learn better.

One way to teach machines which cues matter would be to let them learn on their own and to compare how relevant each cue is in a situation to a human decision maker. Another way would be simply to have the architect of the system explicitly code weights for different cues in different situations that tell the computer how much importance to place on that particular part. Might that raise additional bias issues if one person chooses how important race should be for an algorithm with significant consequences in many peoples’ lives? This question implicates the ethics of machine learning, which will be addressed briefly in Section III.C.¹⁴⁹

For instance, once a bias has been identified in a machine-learning algorithm, there are ways to correct for it.¹⁵⁰ Post training auditing of algorithms provides an opportunity to fix data inputs and analysis that include the potential for bias. It is an “iterative” process, to borrow industry jargon, that requires human interaction in order to rewrite the algorithm to take out the factors that lead to biased results, before training it again and re-evaluating for bias or other undesirable side effects.¹⁵¹ It is relatively easy, for instance, to remove demographic data that could promote explicit training of the algorithm in a biased manner. However, other data such as zip code and socioeconomic status are also highly correlated with demographic information like race, gender, and age, and they may still build bias into the results shown when the algorithm operates.¹⁵²

C. The Remaining Challenges of Ethics and Fairness

One way to evaluate proposed interventions in artificial intelligence processes and procedures to reduce unfairness is to identify which values we seek to promote. A 2018 commission report identified five values for the California state government to consider: autonomy, responsibility, privacy,

149. *See infra* Section III.C.

150. *See* Sunil Madhu, *Are Machines Doomed To Inherit Human Biases?*, FORBES (Aug. 31, 2018, 7:00 AM), <https://www.forbes.com/sites/forbestechcouncil/2018/08/31/are-machines-doomed-to-inherit-human-biases/#7c2a9c01714f> (“When biases are discovered, it is possible to rectify the bias by exposing the machine to more fresh data, feature engineering, algorithm selection, hyperparameter optimization and retraining the machine to eliminate the biased outcome.”).

151. *See* Lindsey Zuloaga, *The Potential of AI to Overcome Human Biases, Rather than Strengthen Them*, HIRE VUE (May 31, 2018), <https://www.hirevue.com/blog/the-potential-of-ai-to-overcome-human-biases-rather-than-strengthen-them>.

152. *See id.*

transparency, and accountability.¹⁵³ Many of the challenges that lawyers face with the use of AI involve issues of fairness; and for many lawyers, fairness is one of the most important, if not the most important, values for client representation. Balancing these interests and values will help legislators and regulators determine the appropriate rules governing uses and abuses of artificial intelligence.

Creating machines that have ethical agency, while challenging, may be on the horizon.¹⁵⁴ Some researchers suggest that “it may be possible to incorporate an explicit ethical component into a machine.”¹⁵⁵ Others note that the machines’ (current) inability to obtain subjective experience means that they cannot learn moral judgments.¹⁵⁶ Challenges remain as to whether that machine could function autonomously, and the most important point for the present is to encourage conversations between those who design these machines and those who specialize in ethics.¹⁵⁷ For more on machines and ethical agency, see Professor Wendel’s article in this symposium issue.¹⁵⁸

An over-arching issue with fairness and AI, however, is that fairness thus far cannot be defined adequately in mathematical terms.¹⁵⁹ Incorporating

153. LITTLE HOOVER COMM’N, ARTIFICIAL INTELLIGENCE: A ROADMAP FOR CALIFORNIA 14 (2018), <https://lhc.ca.gov/sites/lhc.ca.gov/files/Reports/245/Report245.pdf>.

154. See generally Steve Torrance, *Ethics and Consciousness in Artificial Agents*, 22 AI & SOC’Y 495 (2008) (making a case for the future of ethics imbedded in machines).

155. Anderson & Anderson, *supra* note 127, at 25.

156. See, e.g., Joshua P. Davis, *Artificial Wisdom? A Potential Limit on AI in Law (and Elsewhere)*, 72 OKLA. L. REV. 51, 55 (2019). Professor Davis concludes that because “science cannot fully capture the first-person perspective,” which may be necessary “to make moral judgments, and that legal and judicial practitioners sometimes must make moral judgments,” there are limits on AI in legal practice. *Id.* at 88.

157. Anderson & Anderson, *supra* note 127, at 25.

158. See Wendel, *supra* note 54, at 29-35. Professor Wendel explains that “[t]he field of affective computing is in its infancy, and significant progress in this discipline may be required before an AI system can attain a competency required to be an artificial moral agent.” *Id.* at 34-35 (footnote deleted).

159. See SAM CORBETT-DAVIES & SHARAD GOEL, THE MEASURE AND MISMEASURE OF FAIRNESS: A CRITICAL REVIEW OF FAIR MACHINE LEARNING 9-17 (2018), <https://arxiv.org/pdf/1808.00023.pdf> (showing that the three most prominent definitions of fairness—anti-classification, classification parity, and calibration—“suffer from significant statistical limitations”).

Requiring anti-classification or classification parity can, perversely, harm the very groups they were designed to protect; and calibration, though generally desirable, provides little guarantee that decisions are equitable. In contrast to these formal fairness criteria, we argue that it is often preferable to treat similarly risky people similarly, based on the most statistically accurate estimates of risk that one can produce. Such a strategy, while not universally

fairness into the machine-learning rubric, and using human feedback to give instructions about which associations are appropriate and which are not is one potential solution.¹⁶⁰ But fairness is not mathematically quantifiable, as discussed in Part IV, *infra*. Some suggest that it is important for human actors to be present *at every stage of the process* to ask questions to stimulate critical thinking.¹⁶¹

As mathematics is the foundation for AI, the inability to quantify fairness remains a challenge. For instance, should the machine emphasize group fairness or individual fairness? Group fairness might help to reduce subordination, but individual fairness would help to minimize arbitrary classifications.¹⁶² In other words, if we think of fairness as equal outcomes for various groups, the algorithms would need to be calibrated so that those who are less well represented in the data (minorities) will be evaluated differently than those who are more well represented in the data.

On the other hand, if our conception of fairness means that we treat people as individuals rather than as members of groups, under the anti-classification principle, then the algorithms need to produce the same result for people, regardless of their group membership.¹⁶³ Studies show that it is

applicable, often aligns well with policy objectives; notably, this strategy will typically violate both anti-classification and classification parity.

Id. at 1.

160. See Caliskan, Bryson & Narayanan, *supra* note 133, at 183-85.

161. See Elizabeth Isele, *The Human Factor Is Essential to Eliminating Bias in Artificial Intelligence*, CHATHAM HOUSE (Aug. 14, 2018), <https://www.chathamhouse.org/expert/comment/human-factor-essential-eliminating-bias-artificial-intelligence>.

Human agents must question each stage of the process, and every question requires the perspective of a diverse, cost-disciplinary team, representing both the public and private sectors and inclusive of race, gender, culture, education, age and socioeconomic status to audit and monitor the system and what it generates. They don't need to know the answers — just how to ask the questions. In some ways, 21st century machine learning needs to circle back to the ancient Socratic method of learning based on asking and answering questions to stimulate critical thinking, draw out ideas and challenge underlying presumptions. Developers should understand that this scrutiny and reformulation helps them clean identified biases from their training data, run ongoing simulations based on empirical evidence and fine tune their algorithms accordingly. This human audit would strengthen the reliability and accountability of AI and ultimately people's trust in it.

Id.

162. Woodrow Hartzog, *On Questioning Automation*, 48 CUMB. L. REV. 1, 5 (2017) (citing Mark McCarthy, *Standards of Fairness for Disparate Impact Assessment of Big Data Algorithms*, 48 CUMB. L. REV. 67 (2017)).

163. McCarthy, *supra* note 162, at 88.

mathematically impossible to satisfy these notions of fairness together when the underlying data (sentencing records, for example, which are provably harsher on black defendants than on white ones) is imbalanced.¹⁶⁴

Resolving which conceptions of fairness to pursue, and how to do so with AI, is beyond the scope of this Article, and requires continued collaboration across the disciplines. In the meantime, here are four suggestions for improving fairness in lawyers' uses of AI: (1) add diversity to the data sets, (2) diversify technology and design teams, (3) ensure that lawyers have a working knowledge of what AI can and cannot do, and (4) promote more collaboration between lawyers and engineers to augment machine learning to maximize outcomes in justice, if not fairness.

On the issue of diversifying data sets, more careful supervised machine learning can guide AI to make fairer decisions.¹⁶⁵ For instance, there are biases in our language and how we use it, particularly on the internet, so instead of turning bots loose on the internet to learn on their own, a better solution would be to develop a way to teach AI what is acceptable and unacceptable.¹⁶⁶ There is a consortium of artificial intelligence giants, including Amazon, Google, Microsoft, Facebook, and Apple, who participate in the Partnership on Artificial Intelligence to Benefit People and Society, which may be a place to encourage this type of work.¹⁶⁷ The 2018 California Commission report recognized also that the state currently collects insufficient data to do a proper assessment that would more accurately predict the impact of AI on California's future workforce.¹⁶⁸ Collecting that data could be another step in providing an expanded, diverse database for both supervised and unsupervised machine learning.

Diversifying the data sets is closely related to diversifying the teams that design, create, and train AI. In 2016, then-President Obama commissioned a study on the future of artificial intelligence.¹⁶⁹ The study provided some

164. Jon Kleinberg, Sendhil Mullainathan & Manish Raghavan, *Inherent Trade-offs in the Fair Determination of Risk Scores*, in 8TH INNOVATIONS IN THEORETICAL COMPUTER SCIENCE CONFERENCE (ITCS 2017) 43:1 (Christos H. Papadimitriou ed., 2017).

165. See Miro Dudík, John Langford, Hanna Wallach & Alekh Agarwal, *Machine Learning For Fair Decisions*, MICROSOFT: RES. BLOG (July 17, 2018), <https://www.microsoft.com/en-us/research/blog/machine-learning-for-fair-decisions/>.

166. Chris Wiltz, *Bias In, Bias Out: How AI Can Become Racist*, DESIGNNEWS (Sept. 29, 2017), <https://www.designnews.com/content/bias-bias-out-how-ai-can-become-racist/176888957257555>.

167. *Id.*

168. LITTLE HOOVER COMM'N, *supra* note 153.

169. See COMM. ON TECH., NAT'L SCI. & TECH. COUNCIL, EXEC. OFFICE OF THE PRESIDENT, PREPARING FOR THE FUTURE OF ARTIFICIAL INTELLIGENCE (2016),

interesting context on the diversity challenge in artificial intelligence, noting for instance that women constituted only 18% of computer science graduates and about 13% of the participants in the largest AI conference.¹⁷⁰ These numbers led one company to analyze engineering job postings and their language to calculate a gender bias score. It found that in the artificial intelligence sector, the gender bias score was more than twice as high in favor of the male gender, than in any other industry.¹⁷¹

Diversifying the demographics of technology design teams can make a difference in several ways.¹⁷² For instance, those who design the technology devices often do so based on what is important in their worldview (such as “mobile assistants understand voice commands like ‘I’m having a heart attack,’ a health crisis plaguing mostly men, but not ‘I’ve been raped,’ a trauma more likely to fall on a woman”) or what fits in their lives and what fits in their (male) pockets.¹⁷³ Recognizing that the challenge also applies to racial and ethnic minority groups, the President’s working group also recommended commissioning a study “on the AI workforce pipeline in order to develop actions to ensure an appropriate increase in the size, quality, and diversity of the workforce, including AI researchers, specialists, and users.”¹⁷⁴

With more diverse teams, it may be easier to implement the notion of “inclusive intelligence” as it relates to artificial intelligence. Inclusive intelligence integrates both AI research and notions of politics of inclusion.¹⁷⁵ One way to accomplish this goal is to consider how AI disparately impacts underserved communities and the unequal access and disadvantages that it can perpetuate. A lecturer from Columbia suggested “the Four Ds” as ways

https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf.

170. *Id.* at 27-28.

171. *Id.* at 28.

172. See Ari Ezra Waldman, *Designing Without Privacy*, 55 HOUS. L. REV. 659 (2018); see also Melissa Lamson, *Diversity Is Key to Eradicate Bias in AI Solutions*, INC. (Aug. 9, 2018), <https://www.inc.com/melissa-lamson/implicit-bias-can-impact-ai-in-learning-development.html> (“[N]ot only is [sic] essential to diverse teams (of humans) work well together to develop those algorithms—it is imperative that we continue discussing how to manage the potential for problems caused by stereotypes and unconscious biases.”).

173. Waldman, *supra* note 172, at 700.

174. COMM. ON TECH., *supra* note 169, at 28.

175. David Talbot, Levin Kim, Elena Goldstein & Jenna Sherman, *Charting a Roadmap to Ensure Artificial Intelligence (AI) Benefits All*, MEDIUM (Nov. 30, 2017), <https://medium.com/berkman-klein-center/charting-a-roadmap-to-ensure-artificial-intelligence-ai-benefits-all-e322f23f8b59>.

to help make AI more inclusive: “Develop. Decipher. De-identify. De-bias.”¹⁷⁶ “Develop” refers to providing more education about AI to individuals.¹⁷⁷ “Decipher” involves translating computer and engineering language into understandable language for other humans.¹⁷⁸ “De-identifying” data protects the privacy interest by removing personal details where possible.¹⁷⁹ And “De-biasing” involves working to “ensure fairness and avoid digital discrimination,” such as when a search for “baby” produced mostly white infants when using a particular web browser in a country whose population was 64% black.¹⁸⁰

Conclusion

For lawyers specifically, understanding the scope and limits of AI, so that they can properly use, supervise, and evaluate the assistance that it provides, is another crucial component of ensuring fairness. The American Bar Association has a new recommendation regarding technical competence for continuing education certification.¹⁸¹ Will it become a violation of the duty of a lawyer to provide competent representation if she cannot operate in an artificial intelligence world? Does the lawyer need to become an expert in artificial intelligence? How much technological knowledge and ability is going to be expected of attorneys to meet their obligations to be thorough and prepared under the ethical rules and guidelines?¹⁸² In addition, public policy issues are something that computers simply do not understand and are not (thus far) capable of understanding. Lawyers need to be aware of this

176. *Id.*

177. *Id.*

178. *Id.*

179. *Id.*

180. *Id.*

181. Laura van Wyngaarden, *Do Lawyers Have an Ethical Responsibility to Use AI?*, ATT’Y AT WORK (Apr. 5, 2018), <https://www.attorneyatwork.com/do-you-have-an-ethical-responsibility-to-use-ai-in-your-law-practice/>; see Bernstein, *supra* note 78, at 127-28, 132-41 (discussing how to use AI to “mind the gap” of attorneys’ technical competence applying the Model Rules of Professional Conduct).

182. See Roy D. Simon, *Artificial Intelligence, Real Ethics*, N.Y. ST. BAR ASS’N, http://www.nysba.org/Journal/2018/Apr/Artificial_Intelligence,_Real_Ethics/ (last visited May 10, 2019) (cautioning about the duty to supervise nonlawyers and recommending hiring an expert, learning the limits of what AI can do, and double checking the AI product’s output).

limitation, as it will have an impact on conclusions when the machines attempt to reconcile how and why a court would decide a particular way.¹⁸³

More questions come to mind. Will lawyers ethically be required to use AI to save time and money for their clients? Will it be considered a misuse or misallocation of client resources when lawyers summarize documents or do tasks that machines can do more effectively and more efficiently? For some clients who are unable to afford attorneys, legal technology may be their only access to the expertise of a lawyer.¹⁸⁴

Lawyers also need to be cautious of what the profession loses by increasing reliance upon artificial intelligence, and “the question of how organizations adapt over time to domains where human knowledge may erode — or mass cognition may change — because of delegation to AIs.”¹⁸⁵ For instance, the way humans learn or develop institutional knowledge will change when the computers have more of that knowledge, and “[t]hat disruption would affect many organizations’ ability to enhance performance over time, particularly if their reliance on AI were ever disrupted, and to reflect carefully on what is and is not working well.”¹⁸⁶

Transparency is a part of that evaluation process. For instance, the United Kingdom provides an example of protections against bias in machine-driven decisions by requiring companies and government offices to disclose whether a machine has made a decision. If a machine is entirely responsible for making a decision, then that decision can be challenged.¹⁸⁷ Though there are loopholes in this requirement, it still is a step forward.¹⁸⁸

In summary, the challenges for lawyers’ use of AI are similar to those in other fields, and the solutions include diversifying data sets and design teams,

183. See also Becerra, *supra* note 29, at 50 (“Supreme Court cases are often based on policy, and overturned on such. How would AI deal with this overturning of legal precedent that is based on policy? . . . How would a computer be programmed to determine what a reasonable person would do to choose between the charges of second-degree murder or voluntary manslaughter? . . . Just because one has a legal right to do something does not mean it should be done.”).

184. Rostain, *supra* note 80, at 568–69 (“In the public interest and commercial areas, lawyers and technologists have joined forces to address the needs of individuals with no alternative means to solve their legal problems. For most individuals, the choice is not between a technology and a lawyer. It is the choice between relying on legal technologies or nothing at all.”).

185. Cuéllar, *supra* note 30, at 40.

186. *Id.* at 41.

187. Stephen Buryani, *Rise of the Racist Robots – How AI Is Learning All Our Worst Impulses*, GUARDIAN (Aug. 8, 2017, 2:00 PM), <https://www.theguardian.com/inequality/2017/aug/08/rise-of-the-racist-robots-how-ai-is-learning-all-our-worst-impulses>.

188. *Id.*

as well as increasing lawyer competence with technology. Most important, however, is taking steps to promote fairness by de-biasing data, retraining algorithms, and using human actors in supervised machine learning situations to produce fairer outcomes. As Justice Cuéllar notes, “The sooner we realize that we are often already essentially working in organizations of mixed machine and human intelligence, the better we will be to think about the right uses for emerging AI innovations.”¹⁸⁹

189. Cuéllar, *supra* note 30, at 41.