10 things judges should know about AI

BY JEFF WARD

With recent and dramatic advances in the capacities of machine learning, we are now beginning to see artificial intelligence (AI) tools come into their own. This matters for our judiciary, not only because the courts are embedded in an increasingly AI-rich world, but also because AI tools are beginning to enter the courthouse doors, leading to important questions like: Who is liable when an AI tool leads a doctor to a wrong diagnosis? How do defamation laws apply to AI-generated speech? What ground rules should be in place as we use AI tools to assist sentencing? What do hyper-realistic fake videos mean for the rules of evidence?

As AI is rapidly developed and deployed, some rough seas will stir. For all its promise, AI also will challenge our most fundamental commitments to fairness and due process and even our understandings of truth. And, as always, the judiciary will play an essential role in guiding our ships. To that end, here are 10 basic things about AI that every judge should know.

VOLUME 103 NUMBER 1 SPRING 2019

Judicature

Published by the Bolch Judicial Institute at Duke Law. Reprinted with permission. © 2018 Duke University School of Law. All rights reserved.
judicialstudies.duke.edu/judicature

1 AI IS NOT MAGIC. ITS DESIGN, DEVELOPMENT, AND DEPLOYMENT ARE CONSTITUTED IN WAYS THAT REMAIN OUR DOMAIN.

When I turn on my dishwasher, I occasionally stop to think how much time and effort this machine saves me. We are well accustomed to machines replacing our physical efforts. Just as past machines augmented or replaced tasks of human *muscles*, modern AI augments or replaces tasks of the human *mind*. At its heart, AI is the use of machine-based tools to accomplish tasks that would normally require human intelligence.

But AI is not magic. Quite to the contrary, all AI tools are the product of intentional human design, woven from long-held understandings of statistical methodologies, the growing insights of computer science, and the human mind.[1] What has empowered AI's recent advances is neither spell nor sorcery but rather the increasing availability of massive amounts of data and powerful computer processing built to handle that data. These processes are designed and set in motion by human hands.

Why is this important?

The most important message about AI is this: The future paths of AI are neither pre-determined nor beyond our influence. Therefore, engagement from a broad range of stakeholders is essential to walk forward on steady legs. If we treat AI as magic, we may inadvertently cede responsibility and agency to tech companies and limit the roles that other stakeholders can play. AI's design, development, and deployment are constituted, not conjured, and that constitution remains our domain.

2 CURRENT AI TOOLS ARE EFFECTIVE FOR NARROW USES, AND WE CHOOSE IF, WHEN, AND HOW SUCH TOOLS ARE DEPLOYED IN PARTICULAR CONTEXTS.

Current AI is neither C-3PO of *Star Wars* nor HAL 9000 of *2001: A Space Odyssey*. Rather than these human-like, general intelligence agents, state-of-the-art AI is used quite narrowly.[2] Just as my dishwasher is good at a specific task and could neither drive me to work nor mow my lawn, present-day AI tools are made for specific applications, like playing a board game or identifying images indicative of diabetic retinopathy. But they are nowhere near the general and broad intelligences that we know from movies and literature.

Why is this important?

C-3PO could choose to apply its intelligence to nearly any endeavor. Current AI requires *us* to choose the endeavors to which it is applied, so we not only build these tools but also choose if, when, and how they are deployed. Whenever we deploy AI, we are matching a particular *machine* to a particular *context*. This gives us important levers of control and concomitant responsibility to use them. A tool with outputs that are high in accuracy but low in transparency and explainability might be appropriate for autonomous vehicles — where accuracy is preeminent — but may be inappropriate for evaluating due process issues — where articulated reasoning is an inextricable part of our conception of fairness. Therefore, even if a particular AI tool is highly accurate, we might decide not to use it in due process contexts if it suffers from opacity or inexplicability. Alternatively, we might direct the development of a tool in ways that bolster explainability, even at the cost of some measure of accuracy. These are the kinds of engaged choices we must be ready to make, and they demonstrate why this endeavor demands experts from every domain, including the law. For now, we clearly remain the masters of these decisions, choosing if, when, and how such tools are deployed in particular contexts.

▶

3 AI'S RAPID GAINS ARE JUST BEGINNING. WITHOUT ENGAGEMENT, WE RISK BEING CAUGHT OFF GUARD AND MISSING OUR CHANCES TO SHAPE AI'S INFLUENCE.

AI tools are powerful in these narrower settings. We are only just beginning to see substantial progress, widespread use, and the resulting recognition of market opportunities that draw significant money and talent to AI-development endeavors. An interconnected digital world with billions of users has created massive troves of data — long the missing piece in effective AI development — with which to train effective AI tools. If we think of AI development in terms of a growth curve, we are at the start of an arc just beginning to curve exponentially upward.[3]

Why is this important?
The availabilities of big data and strong computing power have fired the starter's gun, and the AI race is on. Its fast pace can be difficult for traditional policy tools and legal rules to track, sometimes catching us off guard and influencing important institutions before we are ready. Without engagement and vigilance, we risk ceding leadership to industry players who may not prioritize societal values among business concerns.

4 MANY OF OUR EFFECTIVE AI TOOLS ARE "DEEP" MACHINE-LEARNING SYSTEMS. THEIR OPACITY MAY CHALLENGE OUR ABILITIES TO EXPLAIN HOW THEY WORK AND TO ENSURE FAIR OUTPUTS.

Symbolic systems of AI are shaped to mimic the logic of humans or experts in a field and, accordingly, compute outputs based on handcrafted sets of rules, logic, and symbols that largely mimic human knowledge. Modern machine-learning systems of AI, by contrast, are designed to learn from data on their own.[4] The former would try to define a "cat" with a symbolic approach of logical rules — if it is furry and has four legs, then it is a cat — while the latter would effectively "teach" the machine what a "cat" is by showing it thousands of example images from which it would discern the best distinguishing features of a cat on its own. Because machine-learning systems determine the decision-making features themselves rather than act upon predetermined rules, their operations can be unintuitive to human minds. The most effective machine-learning tools, however, are often "deep" — complex and multi-layered — models, adding dimensions to the computations that further obscure the processes by which the machine translates data into usable information, even for those who designed those tools.

Why is this important?
The fact that machine-learning AI learns on its own, translating inputs to outputs in ways that are unintuitive to humans, highlights an essential tension: AI may gain its accuracy through some lack of transparency. Deployed carelessly in contexts where transparency matters, such tools may challenge our fundamental need to know, explain, and ensure fairness. Without knowing how an AI tool creates its information outputs, we might not be able to provide the explanations that those affected by these outputs need or deserve. Furthermore, it may be impossible to ensure that the AI tool has not perpetuated unwanted or unlawful biases from its input data. For example, if predominantly orange Tabby cats are provided as input examples, the system may not be able to recognize a Siamese as a cat. And even if the developers of the tool have taken measures at the outset to rid the tool of biases, machine-learning tools may acquire such biases from "learning" in ways that are difficult to spot.

5

MODERN AI CAN GENERATE INFORMATION TO INFORM OUR DECISIONS OR TO MANUFACTURE MEDIA. BOTH FUNCTIONS PRESENT CHALLENGES FOR THE JUDICIARY.

AI tools have two general kinds of outputs: information to guide decisions and synthetic media, such as fake digital photos or videos.

Why is this important? The vast majority of AI tools generate information that helps to make decisions. Here, the output of the AI tools are classifications, categorizations, or probabilities. For example:

- Is the object entering the road a person, another vehicle, or just a shadow?
- Was the word spoken into the search engine “their” or “there”?
- What is the statistical probability that this particular defendant will commit another crime?

With each of these problems, the machine will return a probability or set of probabilities that aims to translate the inputs into information that informs decisions — the decisions of cars, of search engines, of judges, etc.

6 & 7

WHEN USED **APPROPRIATELY**, AI CAN INCREASE THE EFFICIENCY OF OUR PROCESSES AND RAISE THE QUALITY, ACCURACY, AND CONSISTENCY OF OUR DECISIONS.

WHEN USED **INAPPROPRIATELY**, AI CAN UNDERMINE OUR NEED TO KNOW, EXPLAIN, AND ENSURE FAIRNESS, POTENTIALLY PERPETUATING BIASES UNDER GUISES OF OBJECTIVITY.

When used in the administration of law specifically, AI tools may be used to mitigate human biases, bolster evidence-based determinations, and perhaps even provide greater transparency in our judicial reasoning. These benefits are no doubt familiar to judges. The recent history of sentencing guidelines, for example, has highlighted the largely algorithmic — or rules-based — function of criminal sentencing and has incorporated sophisticated statistics and econometrics in part to try to achieve the same benefits.[5] Many states have already incorporated AI tools to assist with pre-trial and post-trial matters, such as bail, parole, and sentencing.[6] Again, there is much to be gained *if* these tools are designed, developed, and deployed appropriately.

On the other hand, recent public discourse has highlighted the potential pitfalls of AI in the courtroom. The AI tool used as part of a pre-sentencing investigation report in *Wisconsin v. Loomis*[7] generated substantial academic and journalistic scrutiny on possible bias in the tool’s computations of inputted data.[8] Despite these concerns, the court found that the defendant’s right to due process was not violated. *Loomis* underscored issues of AI transparency in due process contexts. The plaintiff questioned whether and how gender was weighted in the algorithm’s calculations of criminogenic factors, but the statistical methodologies of the AI tool were never disclosed to the defendant or the court.[9]

The plaintiff in *Loomis* was not the first to question the use of algorithms at sentencing: A line of cases out of Indiana challenged Level of Service Inventory-Revised (LSI-R) scores used as aggravating factors in sentencing decisions.[10] With many states using such tools, these issues are certain to remain relevant. And it will remain essential for judicial institutions to emphasize their gatekeeping functions in order to ensure their own processes are consistent with our expectations of due process and equal protection.

In addition to wrestling with the use of AI in the courts themselves, judges also will increasingly hear cases that contain substantive AI issues on their ▸

facts. For instance, plaintiffs in *DeHoyos v. Allstate Corp.* sought to represent a class of minority policy-holders who claimed that the credit-scoring algorithm used by Allstate-affiliated companies discriminated against minorities in violation of civil rights and federal law.[11] As part of its settlement, Allstate pledged to deploy a revised algorithm to be made publicly available. More recently, Amazon found that an AI tool it was developing to evaluate candidates for tech- and software-related employment positions was not gender-neutral.[12] Because this machine-learning tool was trained on data from a previous decade of successful tech applicants, it learned — perhaps unsurprisingly for a field long dominated by men — that "male" was a marker of a good candidate. Amazon scrapped the tool before applicants' rights were implicated.

Big data sets and the machine-learning tools they feed will have implications in many other contexts as AI tools expand throughout society — from finance to security and defense to medicine. AI's promise to improve medical diagnostics and medical decision-making, for instance, is already a reality in many practice areas. Whether to identify cardiovascular abnormalities or diabetic retinopathies or to direct treatments by better predicting patient responses to drugs, AI tools promise to enhance the accuracy and effectiveness of health interventions. Such use of AI tools raises new questions, such as: When might liability for a misdiagnosis shift from a human doctor to the diagnostic AI tools on which she relied? As industries rely increasingly on machine-based outputs, the consequences of that reliance will become key issues in the courtroom. These issues challenge core legal structures around agency, responsibility, negligence, and malpractice.[13] And deeper challenges will emerge as we work to ensure that opaque machine-learning tools do not undermine our need to know, explain, and ensure fairness.

8

GENERATIVE AI TOOLS — THOSE THAT CREATE SYNTHETIC DATA AND MEDIA — OFFER SIGNIFICANT PROMISE FOR A WIDE RANGE OF CONSUMER AND RESEARCH APPLICATIONS.

In addition to the challenges wrought by machines that inform decisions, the potential promise and peril of AI also grow with advanced machines that make fake digital media. Digital artifacts — fake photos, videos, and voices — made by emerging AI models can fool even the savviest critics. Such fakes are often made from "generative adversarial networks" that pit two AI models against one another: The "generator" model generates digital artifacts intended to fool the "discriminator" model into thinking the generated object is a real example. The outputs of these sophisticated tools may reshape our perceptions of reality and our abilities to trust.[14]

The implications of these hyper-realistic creations are significant both in and out of the courtroom. On the consumer end, these tools may generate content for immersive video games or enhance the resolution of a favorite family photo.[15] Perhaps more profoundly, generative AI is poised to enhance medical research: For example, in settings where very limited real data is available, generative AI might be used to create additional, realistic examples to augment data sets that, in turn, train the tools that will bring us more accurate and more accessible diagnostics.[16] Generative AI is already empowering astronomers and other scientists by reconstructing data originally transmitted at low-resolutions, building rich and usable data for scientific study.[17] This is only the beginning.

9

WHEN USED TO CREATE FAKE MEDIA, GENERATIVE AI MAY THREATEN OUR ABILITIES TO TRUST AND TO DISCERN REALITY, POSING CHALLENGES TO FUNDAMENTAL CIVIC INSTITUTIONS AND PROCESSES.

Once again, though, this promising technology poses significant challenges. Generative AI is rather new, but it is not difficult to imagine the problems that synthetic, hyper-realistic digital artifacts could create for courtroom adjudication. Fake digital artifacts might include a photograph showing the defendant present at the scene of the assault; a video recording indicating the property was already damaged before the time of the accident; a voice recording that sounds like the CEO unlawfully conspiring. As the technology advances, anyone with a smart phone may have the ability to make the unreal seem real and to force us to question our sense of what can be trusted.

What happens when current rules of evidence do not keep pace with these advances? As just one timely example, Federal Rule of Evidence 902 was recently amended to increase the list of "items of evidence that are self-authenticating" and "require no extrinsic evidence of authenticity in order to be admitted."[18] For all its benefits, making it easier to authenticate digital evidence may prove problematic when hyper-realistic fakes generated by sophisticated AI tools become more prevalent. By what standards will such items need to be authenticated? And what must we do to ensure that the evidence we rely on is true?[19]

Of course, if these technologies can fool qualified and knowledgeable people and even experts in a courtroom, we must examine their potential for broader public effects. A fake (but seemingly real) video showing the president announcing military action by a hostile state, for example, may incite public unrest or create national security issues.[20] And as damaging as any isolated use of such technology may be, the ubiquitous use of hyper-realistic fakes could also threaten something even more fundamental — our ability to trust public discourse and democratic institutions.[21]

10

THE JUDICIARY — AMONG MANY OTHER STAKEHOLDERS — MUST PLAY A KEY ROLE IN ENSURING THAT AI TOOLS DO NOT UNDERMINE OUR CORE CULTURAL VALUES.

To be sure, AI promises both to improve our lives and to challenge our most fundamental conceptions of fairness, due process, and even truth. And, perhaps ironically, technology itself will play a role in helping us to combat these technological dangers — technology, for example, may be able to audit the algorithms used in sentencing or to help us spot fake digital creations.[22] But with so much at stake, we cannot rely on technology alone, and the judiciary — among many other stakeholders — will be called upon to play a key role in ensuring that AI tools do not undermine our core cultural values. There is cause for optimism: The common law is a system built for evolution, and the judiciary has proven adept at learning new worlds and helping to guide our ships through uncharted seas. ▶

[1] Keith McNulty, *What is Machine Learning?*, Machine: Towards Data Science (Aug. 7, 2018), https://towardsdatascience.com/what-is-machine-learning-891f23e848da.

[2] Michale Chalfen, *The Challenges of Building AI Apps*, TechCrunch (2015), https://techcrunch.com/2015/10/15/machine-learning-its-the-hard-problems-that-are-valuable/.

[3] Yoav Shoham et al., Stanford University, The AI Index 2018 Annual Report (2018), http://cdn.aiindex.org/2018/AI%20Index%202018%20Annual%20Report.pdf.

[4] *Symbolic Reasoning (Symbolic AI) and Machine Learning*, Skymind: Artificial Intelligence Wiki, https://skymind.ai/wiki/symbolic-reasoning (last visited Jan. 8, 2019).

[5] Matthew Van Meter, *One Judge Makes the Case for Judgment*, The Atlantic (Feb. 25, 2016), https://www.theatlantic.com/politics/archive/2016/02/one-judge-makes-the-case-for-judgment/463380/ ("The sentencing guidelines are essentially an algorithm. For each charge, the judge inputs the crime's 'base offense level' and makes adjustments based on factors like the defendant's role in the crime, his acceptance of responsibility, and whether he had a gun. The resulting value, from one to 43, is the defendant's 'offense level.' The judge then uses a table to cross-reference the offense level with another number based on the defendant's prior convictions. At the intersection of the offense level and the criminal-history category, the judge finds the 'guideline range,' an upper and lower sentencing limit. Some judges call this process 'grid and bear it.'").

[6] *Algorithms in the Criminal Justice System*, EPIC, https://epic.org/algorithmic-transparency/crim-justice/ (last visited Jan. 8, 2019).

[7] *Wisconsin v. Loomis*, 881 N.W.2d 749 (Wis. 2016).

[8] *See, e.g.*, Julia Angwin et al., *Machine Bias*, ProPublica (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (discussing software used to predict future criminals that is biased against blacks); Elaine Angelino et al., *Learning Certifiably Optimal Rule Lists for Categorial Data*, 18 J. Machine Learning Research 1 (2018), https://app.scholasticahq.com/supporting_files/2110842/attachment_versions/2115482.

[9] *Loomis*, 881 N.W.2d at 765.

[10] *See Rhodes v. Indiana*, 896 N.E.2d 1193, 1196 (Ind. Ct. App. 2008) (holding trial court abused its discretion by relying on defendant's LSI-R as an aggravating sentencing factor); *Rodgers v. Indiana*, No. 79A05-0612-CR-731, 2007 Ind. App. Unpub. LEXIS 706 at *3 (Ind. Ct. App. Aug. 31, 2007) (reviewing trial court's finding of low LSI-R score as a mitigating factor for defendant's sentencing).

[11] *DeHoyos v. Allstate Corp.*, 240 F.R.D. 269 (W.D. Tex. 2007).

[12] Jeffrey Dastin, *Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women*, Reuters (Oct. 9, 2018, 11:12 p.m.), https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G.

[13] *See* Michael Woolf, *"Paging Dr. Bot" – The Emergence of AI and Machine Learning in Healthcare,* ABA (Sept. 27, 2018), https://www.americanbar.org/groups/health_law/publications/aba_health_esource/2016-2017/october2017/machinelearning/; Shailin Thomas, *Artificial Intelligence, Medical Malpractice, and the End of Defensive Medicine*, Harvard Law: Bill of Health (Jan. 26, 2017), http://blog.petrieflom.law.harvard.edu/2017/01/26/artificial-intelligence-medical-malpractice-and-the-end-of-defensive-medicine/ (noting, "But who should be responsible when a doctor provides erroneous care at the suggestion of an AI diagnostic tool? If the algorithm has a higher accuracy rate than the average doctor — as many soon will — it seems wrong to continue to place blame on the physician. Going with the algorithm's suggestion would always be statistically the best option — so it's hard to argue that a physician would be negligent in following the algorithm, even if it turns out to be wrong and the doctor ends up harming a patient. As algorithms improve and doctors use them more for diagnosing and decision-making, the traditional malpractice notions of physician negligence and recklessness may become harder to apply.").

[14] Ganes Kesari, *Catch Me if You Can: A Simple English Explanation of GANs or Dueling Neural-nets*, Medium: Towards Data Science (Mar. 28, 2018), https://towardsdatascience.com/catch-me-if-you-can-a-simple-english-explanation-of-gans-or-dueling-neural-nets-319a273434db.

[15] Heather Alexandra, *A Look At How No Man's Sky's Procedural Generation Works*, Kotaku (Oct. 18, 2016, 12:30 p.m.), https://kotaku.com/a-look-at-how-no-mans-skys-procedural-generation-works-1787928446.

[16] Michal Amitai et al., *Synthetic Data Augmentation Using GAN for Improved Liver Lesion Classification*, arXiv preprint arXiv:1801.02385v1 (Jan. 8, 2018), https://arxiv.org/pdf/1801.02385.pdf. Note: This is important where the training of AI-driven diagnostic tools is stifled because the privacy-protected medical data needed to train the tools is expensive, time intensive, or simply not available in significant quantities.

[17] Davide Castelvecchi, *Astronomers Explore Uses for AI-Generated Images*, Nature (Feb. 2 2017), https://www.nature.com/polopoly_fs/1.21398!/menu/main/topColumns/topLeftColumn/pdf/542016a.pdf.

[18] Andrew Toft, *New Rules for Self-Authenticating Electronic Evidence*, ABA (June 22, 2018), https://www.americanbar.org/groups/litigation/committees/trial-evidence/practice/2018/new-rules-electronic-evidence/.

[19] *The "Deep Fake" Video Threat*, Bloomberg (June 13, 2018, 7:00 a.m.), https://www.bloomberg.com/opinion/articles/2018-06-13/the-deep-fake-video-threat.

[20] Robert Chesney & Danielle Citron, *Deep Fakes: A Looming Crisis for National Security, Democracy and Privacy?*, LawFare (Feb. 21, 2018, 10:00 a.m.), https://www.lawfareblog.com/deep-fakes-looming-crisis-national-security-democracy-and-privacy; Bruno S. Frey et al., *Will Democracy Survive Big Data and Artificial Intelligence?*, Scientific American (Feb. 25, 2017), https://www.scientificamerican.com/article/will-democracy-survive-big-data-and-artificial-intelligence/.

[21] Miles Brundage et al., *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation*, 43, 46 (Feb. 2018), https://arxiv.org/pdf/1802.07228.pdf ("Worryingly, the features of AI . . . make it particularly well suited to undermining public discourse through the large-scale production of persuasive but false content, and strengthening the hand of authoritarian regimes. . . . Even if bots users only succeed in decreasing trust in online environments, this will create a strategic advantage for political ideologies and groups that thrive in low-trust societies or feel opposed by traditional media channels. Authoritarian regimes in particular may benefit from an information landscape where objective truth becomes devalued and 'truth' is whatever authorities claim it to be.") .

[22] Karen Hao, *Deepfake-busting Apps Can Spot Even a Single Pixel Out of Place*, MIT Technology Review (Nov. 1, 2018), https://www.technologyreview.com/s/612357/deepfake-busting-apps-can-spot-even-a-single-pixel-out-of-place/; Elizabeth Gibney, *The Scientist Who Spots Fake Videos*, Nature (Oct. 6, 2017), https://www.nature.com/news/the-scientist-who-spots-fake-videos-1.22784.

JEFF WARD is Director of the Duke Center on Law & Tech (DCLT), Associate Dean for Technology & Innovation, and Associate Clinical Professor of Law at Duke Law. He teaches courses at the intersection of law and emerging technologies, offers the Duke Law Tech Lab pre-accelerator program for early-stage legal tech companies, and uses the tools of the law to help ensure that new technologies ultimately empower and ennoble people and to expand access to quality legal services.