MACHINE LEARNING EVIDENCE: ADMISSIBILITY AND WEIGHT

*Patrick W. Nutter**

INTRODUCTION

Artificial intelligence ("AI") is gaining traction in legal practice. How prosecutors prioritize which crimes to prosecute,[1] sift through mountains of documents,[2] and establish reasonable suspicion[3] can all reasonably be expected to change with coming AI technologies. While lawyers need not attain expert-level knowledge of these processes, some competency in concepts and vocabulary will be essential, in the same manner it has been with other sciences, like statistical evidence or DNA analysis. In that vein, this Comment aims to give attorneys a much-needed look inside the "black box" of one emerging type of AI technology, machine learning. With at least some familiarity with how machine learning works, attorneys can begin to formulate questions and strategies when that kind of technology produces substantive evidence at trial. These include potential issues under the Fifth and Sixth Amendments as well as the Federal Rules of Evidence, none of which, I argue, would categorically bar machine learning evidence. After establishing that machine learning evidence is admissible, I explain how counsel for both sides must be aware of the significant issues with machine learning that nonetheless could affect the weight such evidence is assigned by the trier of fact.

Machine learning refers to a process in which a "machine has been 'trained' through exposure to a large quantity of data and infers a rule from the patterns it observes."[4] The technology, once only theoretical, is now

* J.D. Candidate, 2019, University of Pennsylvania Law School; B.A., 2015, University of California, Irvine. I would like to thank Professor Jonathan Klick and Professor David Rudovsky for their advice on this Comment. I also thank the dedicated editors of the *University of Pennsylvania Journal of Constitutional Law* for their assistance in bringing this Comment to fruition.

1 *See* Andrew Guthrie Ferguson, *Predictive Prosecution*, 51 WAKE FOREST L. REV. 705, 732 (2016) ("[T]he predictive prosecution model shifts the identification of problem areas from the street cops to the lawyers.").

2 Harry Surden, *Machine Learning and Law*, 89 WASH. L. REV. 87, 110–15 (2014).

3 *See* Stephanie Lacambra, *Predictive Policing: A Guide for Criminal Defense Attorneys*, ELEC. FRONTIER FOUND., https://www.eff.org/files/2017/10/30/predictive_policing_one_pager.pdf (last visited Mar. 7, 2018) (defining and explaining the use of "predictive policing" by law enforcement).

4 Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. PA. L. REV. 633, 679 (2017).

responsible for many tasks in daily digital life. For instance, machine learning is at work when Facebook automatically recognizes a user in a photo[5] or when an email client automatically routes spam to the appropriate folder.[6]

For many litigators, it will only be a matter of time before they first encounter a creative opposing counsel who wishes to admit machine learning output into evidence. When that happens, both sides in the interests of clients—and the court in the interest of the law itself—must be equipped with certain questions and skepticism. This Comment aims to look ahead to possible evidentiary issues when, not if, the output of machine learning algorithms is used as substantive evidence in criminal prosecution.

In the very near future, AI software will affect criminal and civil litigation in at least three significant ways. First, AI will pose the critical question of whether and to what extent the decision of the algorithm exposes the user to liability.[7] For example, in the employment context, when an algorithm pre-screens resumes and, not by intentional design, discounts the resumes of women or minorities, is the employer liable for discrimination?[8] Or, since the technology will soon be deployed on police body cameras,[9] could real-time object recognition software perhaps assist an officer by identifying whether a gun or a smartphone is in the suspect's hand, and what liability might exist if the algorithm decided incorrectly?[10] Second, AI will also alter predictive technologies in the criminal justice system, such as ones that may

---

5 *See* Daniel Terdiman, *Facebook's Image-Recognition Tech Is Teaching 40,000 Images a Second to Understand Context*, FAST CO., (June 8, 2017), https://www.fastcompany.com/40428910/facebooks-image-recognition-tech-is-teaching-40000-images-a-second-to-understand-context ("For [Facebook's] 1.94 billion monthly users, artificial intelligence and machine learning are behind the ability to quickly surface meaningful baby pictures, vacation selfies, and pet action photos.").

6 *See* Surden, *supra* note 2, at 90–93 (discussing email spam filters as an example of machine learning).

7 *See generally* Rebecca J. Krystosek, *The Algorithm Made Me Do It and Other Bad Excuses: Upholding Traditional Liability Principles for Algorithm-caused Harm*, MINN. L. REV. DE NOVO (May 17, 2017), http://www.minnesotalawreview.org/2017/05/the-algorithm-made-me-do-it-and-other-bad-excuses/ (discussing various forms of legal liability for the actions and decisions of algorithms).

8 *See* Hannah Devlin, *AI Programs Exhibit Racial and Gender Biases, Research Reveals*, GUARDIAN (Apr. 13, 2017, 2:00 PM), https://www.theguardian.com/technology/2017/apr/13/ai-programs-exhibit-racist-and-sexist-biases-research-reveals ("One previous study showed that an identical CV is 50% more likely to result in an interview invitation if the candidate's name is European American than if it is African American. The latest results suggest that algorithms, unless explicitly programmed to address this, will be riddled with the same social prejudices.").

9 *See* Drew Harwell, *Facial Recognition May Be Coming to a Police Body Camera Near You*, WASH. POST (Apr. 26, 2018), https://www.washingtonpost.com/news/the-switch/wp/2018/04/26/facial-recognition-may-be-coming-to-a-police-body-camera-near-you/?utm_term=.46192f40bfda (describing the growing use of facial recognition software in police body cameras).

10 *See, e.g.*, Eric Levenson, Madison Park & Darran Simon, *Sacramento Police Shot Man Holding Cellphone in His Grandmother's Yard*, CNN (Mar. 22, 2018) https://edition.cnn.com/2018/03/22/us/sacramento-police-shooting/index.html (reporting a shooting of a man holding a cell phone by an officer who believed the phone to be a gun).

aid in investigations, establish reasonable suspicion or probable cause,[11] or assist sentencing judges in estimating a defendant's chances of reoffending.[12] Third, AI can aid the legal reasoning process itself. For example, to understand the original public meaning of the Second Amendment's "bear arms," it would surely be illuminating to examine a corpus of 1.3 billion words—from books, handwritten diaries, newspapers, etc.—for the use of the phrase "bear arms" in the centuries surrounding the Amendment's drafting, a task that has been accomplished with AI technology.[13]

Despite the important developments and commentary on those evolving issues, this Comment focuses specifically on using the conclusions of machine learning processes as substantive evidence in litigation. For instance, in a blurry surveillance video or an unclear audio recording, the naked eye and ear may be insufficient to prove guilt beyond a reasonable doubt, but certain recognition algorithms could do so easily. Lip-reading algorithms might tell jurors what was said on video where there is no audio available.[14] A machine might construct an estimation of a perpetrator's face from only a DNA sample,[15] or in other DNA analysis of corrupted samples.[16]

---

11 *See generally* Michael L. Rich, *Machine Learning, Automated Suspicion Algorithms, and the Fourth Amendment*, 164 U. PA. L. REV. 871 (2016) (discussing the role of machine learning will play in the Fourth Amendment context).

12 *See* Ellora Thadaney Israni, Opinion, *When an Algorithm Helps Send You to Prison*, N.Y. TIMES (Oct. 26, 2017), https://www.nytimes.com/2017/10/26/opinion/algorithm-compas-sentencing-bias.html (criticizing the use of a risk assessment algorithm as a factor in a criminal sentencing).

13 *See* Johnson, *Arms and the Man*, ECONOMIST (June 9, 2018), http://media.economist.com/news/books-and-arts/21743640-big-data-can-help-clarify-meaning-second-amendment-judges-should-pay (advocating for the use of "digital corpora" to determine the meaning of the term "bear arms").

14 *See* Jamie Condliffe, *AI Has Beaten Humans at Lip-reading*, MIT TECH. REV. (Nov. 21, 2016), https://www.technologyreview.com/s/602949/ai-has-beaten-humans-at-lip-reading/ (describing two studies in which artificial intelligence vastly outperformed humans at lip-reading).

15 The leading company offering this particular service is Virginia-based Parabon Nanolabs, which uses machine learning processes to predict visible traits (e.g., facial structure, eye and hair color, etc.) from DNA samples alone. *See How DNA Phenotyping Works*, PARABON NANOLABS, https://snapshot.parabon-nanolabs.com/#phenotyping-how (last visited Mar. 7, 2018) ("Parabon's scientists use machine learning algorithms to combine the selected set of SNPs into a complex mathematical equation for the genetic architecture of the trait."). Parabon's service has already been used in several investigations. *See, e.g.*, Alicia Victoria Lozano, *Montgomery County Officials Use DNA Samples to Create Picture of Rape Suspect*, NBC PHILA. (Jan. 16, 2018, 3:12 PM), https://www.nbcphiladelphia.com/news/local/Montgomery-County-Phenotyping-DNA-Testing-Rape-Suspect-Norristown-Farm-Park-469588793.html (last updated Jan. 16, 2018, 7:29 PM)(discussing the Montgomery County District Attorney Office's use of Parabon's DNA technology to create an illustration of a suspected rapist). It is, however, not without critics. One, Peter Claes, an expert in craniofacial morphometrics at the University of Leuven, thinks that in some cases the images have virtually no value. To him, one image "just looked like an average black man. It didn't have any characteristic features. That reconstruction didn't give any more information than the genetic background that they listed. This prediction is hardly specific so it doesn't really focus on an individual . . . ." Howard Wolinsky, *CSI on Steroids*, 16 EMBO REP. 782, 782 (2015).

16 *See Under the Microscope—Jonathan Adelman & Michael Marciano*, ISHI (Sept. 21, 2017),

The legal issues of broadly defined "machine evidence" have been extensively cataloged and discussed, especially in the Fourth and Sixth Amendment contexts.[17] Such machine evidence includes radar guns, breathalyzers, DNA analysis software, GPS, and risk assessment software.[18] However, few have explored machine learning as a distinct species of machine evidence, distinct even from evidence produced using traditional computer programs,[19] with its own vocabulary and unique set of issues. Importantly, that lack of analysis means there has been little exploration of the legal pitfalls of machine learning—the ways in which it goes awry, is misused, or is misinterpreted. In some ways, the reliability issues of machine learning algorithms are similar to those already cataloged with respect to typical computer software; but in other critical respects, machine learning poses unique questions of reliability. Like other machine evidence has done in the past, machine learning will give rise to new evidentiary issues. Ultimately, however, I argue that in most cases machine learning evidence will not be barred by either the Federal Rules of Evidence or the Fifth and Sixth Amendments to the Constitution.

Part I begins with an overview of how courts currently treat software output as evidence. Machine learning is revolutionary in its applications and capabilities, though, with respect to its potential uses in prosecution, it is functionally similar to traditional software: data go in and conclusions come out. In between, there is a "black box" of calculations that few in the courtroom understand. Part II explains how machine learning is distinct from traditional computer software in process and appropriate uses. Part III offers an explanation of how contemporary machine learning typically works.

In Part IV, I analyze machine learning evidence under Federal Rule 702 and its *Daubert* criteria and find that machine learning would surely meet the requirements for admissible expert testimony.

In Part V, I argue that the Fifth and Sixth Amendments pose no categorical barrier to machine learning evidence but limit how it may be introduced. I argue first that the Fifth Amendment's Due Process Clause does not bar machine learning evidence and, second, that pursuant to the Sixth Amendment's Confrontation Clause, machine learning evidence will

---

https://www.ishinews.com/under-the-microscope-michael-marciano_jonathan_adelman/ (interviewing two experts on the issues of DNA mixture interpretation).

17 *See generally* Andrea Roth, *Machine Testimony*, 126 YALE L.J. 1972 (2017) [hereinafter Roth, *Testimony*] (attempting to "offer a coherent framework for conceptualizing and regulating machine evidence"); Andrea Roth, *Trial by Machine*, 104 GEO. L.J. 1245 (2016) [hereinafter Roth, *Trial*] (discussing the rise of machines in criminal adjudication).

18 Roth, *Testimony, supra* note 17, at 2015, 2025, 2027.

19 *See* Christian Chessman, Note, *A "Source" of Error: Computer Code, Criminal Defendants, and the Constitution*, 105 CALIF. L. REV. 179, 183–84 (2017) (discussing evidentiary issues with respect to conventional computer software).

likely only be admissible in the form of expert testimony.

In Part VI, having concluded that machine learning evidence will likely be admissible in at least some cases, I emphasize that there are significant problems with the weight such evidence should be assigned by the finder of fact because of machine learning's unique unexplainability, that is, in many cases it is impossible to explain how a machine learning algorithm makes a particular conclusion.

I. MACHINE EVIDENCE AND BLACK BOXES

Evidence is "[s]omething (including testimony, documents, and tangible objects) that tends to prove or disprove the existence of an alleged fact" or, more generally, "anything presented to the senses and offered to prove the existence or nonexistence of a fact."[20] In a criminal proceeding, evidence, and the inferences that logically can be drawn from it, must ultimately support the factfinder's conclusion of guilt.[21] The primary purpose of rules of evidence is to narrow the evidence offered at trial, sometimes to limit evidence to what is relevant and probative, other times to prevent the factfinder from drawing illogical conclusions or to minimize the possibility of unfair prejudice to the accused.[22] Where the Federal Rules of Evidence apply, they explicitly instruct courts to construe them in a manner that will "administer every proceeding fairly, eliminate unjustifiable expense and delay, and promote the development of evidence law, to the end of ascertaining the truth and securing a just determination."[23]

How guilt may be established has evolved over the course of the Anglo-American legal tradition. Whereas documentary evidence and human testimony have been mainstays of criminal proceedings, other forms of evidence have unfortunately come and thankfully gone, including phrenology[24] and "spectral evidence" (i.e., the "testimony of the bewitched

---

20 *Evidence*, BLACK'S LAW DICTIONARY (10th ed. 2014).

21 1 WHARTON'S CRIMINAL EVIDENCE § 1:2 (15th ed. 1997)

22 *Id.*

23 FED. R. EVID. 102.

24 In the latter half of the nineteenth century, Cesare Lambroso formulated and evangelized his own "scientific" classification of criminals and testified often as an expert witness:

> He noticed in the skull of a murderer an anomalous depression characteristic of lower species, such as dogs. . . . Lombroso speculated that such a skull reflected an underlying brain abnormality of an atavistic nature. That is, perhaps the brain of the murderer suggested a more primitive development of a lower species. Lombroso gathered large quantities of data from measurements on criminals and proposed that certain criminals represented a distinct species, *homo delinquens*. As his reputation grew, others also subscribed to his theory that at least some criminals are born, not made, and criminal types could be identified by the shapes of their skulls. Lombroso was called upon as an expert witness on numerous occasions to testify as to whether a defendant was of a criminal disposition.

that an accused person's spectral shape appeared to them at a time when their physical body was elsewhere").[25] "Machine evidence," however, has come and stayed. Over the past 150 years, the "silent testimony of instruments" has supplemented the testimony of humans.[26] Only rarely have courts found that science had progressed too far beyond what the given rules of evidence can comfortably handle and thus resisted a new technology as evidence.[27] Instead, the law has typically been receptive to new scientific discoveries and their potential evidentiary uses.[28]

Overall, "this shift from human- to machine-generated proof has, on the whole, enhanced accuracy and objectivity in fact finding."[29] And yet, for all its advantages, machine testimony is not without risks, such as when society determines that it must err on the side of overinclusion and reduction of false negatives, notwithstanding such a policy's inherent risk that machines will erroneously inculpate the innocent.[30] This Comment highlights that risk, as well as another: that machines are improperly afforded a presumption of reliability, even when jurors cannot peer into the "black box" that is providing them with evidence. "These 'black box' processes, because of their mechanical appearance and apparently simple output, have a veneer of objectivity and certainty."[31] However, even though these machines appear neutral, they are necessarily the product of human creation, and therefore human judgment, with its risk of bias and tendency to make mistakes.[32]

Richard L. Elliott, *Neuropsychiatry in the Courtroom*, 62 MERCER L. REV. 933, 939 (2011).

25 *See* Sarah Kruetter, *The Devil's Specter: Spectral Evidence and the Salem Witchcraft Crisis*, 2 SPECTRUM: A SCHOLARS DAY J., 1, 1 (2011) ("This was a key point of proof delivered against accused witches at Salem in 1692. Spectral evidence is impossible to prove and courts used it with caution in court cases prior to Salem. . . . [Y]et nearly every case during the Salem outbreak featured this evidence.").

26 Roth, *Trial*, *supra* note 17, at 1253 (quoting MIRJAN R. DAMAŠKA, EVIDENCE LAW ADRIFT 143 (1997)); *see id.* ("'[S]cientific gadgets in the law of evidence' and interpretive forensic and diagnostic software has reduced the role of both percipient and human witnesses in proving guilt." (quoting Note, *Scientific Gadgets in the Law of Evidence*, 53 HARV. L. REV. 285, 285 (1939) [hereinafter *Scientific Gadgets*])).

27 *See, e.g.*, People v. Offermann, 125 N.Y.S.2d 179, 185 (N.Y. Sup. Ct. 1953) (explaining that the case was the very first to use a radar gun reading as evidence and holding that the New York legislature should enact new rules of evidence to explicitly allow for its admissibility).

28 *Scientific Gadgets*, *supra* note 26, at 285 ("It is the perennial boast of the law that in the ascertainment of facts it will avail itself of any accepted scientific discovery.").

29 Roth, *Testimony*, *supra* note 17, at 1976.

30 Roth, *Trial*, *supra* note 17, at 1269 ("[M]echanization has arisen in criminal justice in an unbalanced way, reflecting the focus of law enforcement, interest groups, and lawmakers on reducing a particular species of inaccuracy: false negatives.").

31 *Id.* at 1269–70.

32 *Id.* at 1270 ("In truth, these processes all have hidden subjectivities and errors that often go unrecognized and unchecked, thus potentially 'facilitat[ing] the masking of illegitimate or illegal discrimination behind layers upon layers of mirrors and proxies.' (quoting Omer Tene & Jules Polonetsky, *Judged by the Tin Man: Individual Rights in the Age of Big Data*, 11 J. TELECOMM. & HIGH TECH. L. 351, 358 (2013)).

Even so, machine evidence—and, for the purposes of this Comment, specifically evidence derived from algorithmic software processes—supports guilty verdicts daily. By their conduct, courts have expressed a tolerance for some level of both ignorance and risk in machine evidence: ignorance in how these processes work, and risk that they might not "get it right" every time. For example, photographic evidence, breathalyzer readouts, and DNA tests have been admitted into evidence for decades, in spite of their risk of error in programming or hidden reliance on subjective human judgment.[33] In recent cases involving TrueAllele, a probabilistic genotypic software,[34] the black box has only gotten blacker, and courts have yet to reject its use on that basis. Indeed, TrueAllele's most marketable feature is the assumptions it uses to *remove* user (that is, lab technician) judgment from the DNA match determination, effectively promoting its "veneer of objectivity and certainty." This merely passes the buck, however, as the user's judgment is only substituted for that of the initial programmer of the software, who, as of now, has never revealed his complete methodology and has not been subject to cross-examination.[35]

Overall, then, courts have long been comfortable with machine evidence whose processes are not entirely disclosed to, or understood by, the judge, jury, parties, or counsel. And it is likely that courts will find similar comfort in machine learning processes.

II. WHAT MACHINE LEARNING IS

A. Machine Learning in the Artificial Intelligence Context

Technologies that claim the artificial intelligence label are proliferating in number and application. A 2016 Stanford University report lists no fewer than eight broad sectors that researchers are hoping to transform with AI, including critical areas like education, healthcare, transportation, the workplace, and public safety.[36] Yet for many AI researchers, listing even eight sectors is too

---

33 *Id.* at 1272–73 (discussing the potential errors and human judgments that inform how photographs, breathalyzers, and DNA tests operate).

34 Probabilistic genotyping "uses complex mathematical formulas to examine the statistical likelihood that a certain genotype comes from one individual over another." Jessica Pishko, *The Impenetrable Program Transforming How Courts Treat DNA Evidence*, WIRED (Nov. 29, 2017, 7:00 AM), https://www.wired.com/story/trueallele-software-transforming-how-courts-treat-dna-evidence.

35 Roth, *Trial*, *supra* note 17, at 1273–74.

36 COMMITTEE OF THE ONE HUNDRED YEAR STUDY OF ARTIFICIAL INTELLIGENCE, STUDY PANEL, ARTIFICIAL INTELLIGENCE AND LIFE IN 2030, at 4 (2016).

conservative: AI will simply transform everything.[37] To them, that future is "when," not "if." The necessary technologies are already here, but their wider applications are presently constrained only by human imagination, management,[38] and the sheer lack of people working in the field.[39]

AI has infected the discourse of business and culture perhaps because it seems to refer to so many things. Firms increasingly market themselves as incorporating AI into their products and services,[40] though sometimes they use the AI label inaccurately, applying its traditional computational methods only for more marketing heft.[41] More often, the term is used in platitudes about market "disruption,"[42] and in fact "AI" may now be used so loosely that it is losing its meaning—what one Georgia Institute of Technology professor calls "AI deflation."[43] Even as a field of study, artificial intelligence has hazy boundaries, as it refers to many disparate specialties like robotics,[44] transportation,[45] human-computer interaction,[46] and predictive technologies.[47] As such, one researcher comments, "the field doesn't have a

---

37 Erik Brynjolfsson & Andrew McAfee, *The Business of Artificial Intelligence*, HARV. BUS. REV., July 2017 at 3, 4 ("The effects of AI will be magnified in the coming decade, as manufacturing, retailing, transportation, finance, health care, law, advertising, insurance, entertainment, education, and virtually every other industry transform their core processes and business models to take advantage of machine learning.").

38 *Id.* ("The bottleneck now is in management, implementation, and business imagination.").

39 Cade Metz, *Tech Giants Are Paying Huge Salaries for Scarce A.I. Talent*, N.Y. TIMES (Oct. 22, 2017), https://www.nytimes.com/2017/10/22/technology/artificial-intelligence-experts-salaries.html?_r=1 ("In the entire world, fewer than 10,000 people have the skills necessary to tackle serious artificial intelligence research . . . .").

40 *See, e.g.*, Kate Kaye, *Is This AI or BS? Artificial Intelligence Is All the Rage, but Sometimes It's Just Hype*, ADAGE (Apr. 19, 2017), http://adage.com/article/datadriven-marketing/ai-bs/308718/ (discussing the marketing power and oversimplification of the buzzwords "artificial intelligence").

41 Brynjolfsson & McAffee, *supra* note37, at 4 ("Simply calling a dating site 'AI-powered,' for example, doesn't make it any more effective, but it might help with fundraising.").

42 *See e.g.*, *The AI Disruption Bundle: The Guide to Understanding How Artificial Intelligence Is Impacting the World*, BUS. INSIDER (Oct. 6, 2017, 3:30 PM), http://www.businessinsider.com/understanding-artificial-intelligence-impacting-world-2017-10 (describing artificial intelligence as disruptive).

43 Ian Bogost, *'Artificial Intelligence' Has Become Meaningless*, ATLANTIC (Mar. 4, 2017), https://www.theatlantic.com/technology/archive/2017/03/what-is-artificial-intelligence/518547/ (referencing artificial intelligence robots).

44 Felix Ingrand & Mark Ghallab, *Robotics and Artificial Intelligence: A Perspective on Deliberation Functions*, AI COMMUNICATIONS, IOS PRESS (Apr. 3, 2015), https://hal.archives-ouvertes.fr/hal-01138117/document (discussing AI and robotics).

45 *See generally* Sebastian Ramos et al., *Detecting Unexpected Obstacles for Self-Driving Cars: Fusing Deep Learning and Geometric Modeling*, ARXIV.ORG (Dec. 20, 2016), https://arxiv.org/pdf/1612.06573.pdf (discussing AI and cars).

46 *See generally* Jose Maria Garcia-Garcia et al., *Emotional Detection: A Technology Review*, RESEARCHGATE (2017), https://www.researchgate.net/profile/Jose_Garcia-Garcia4/publication/320359659_Emotion_detection_a_technology_review/links/59e620a2a6fdcc3dcd33e82f/Emotion-detection-a-technology-review.pdf (discussing human-AI interactions).

47 *See generally* David Silver et al., *Mastering the Game of Go Without Human Knowledge*, 550 NATURE 356

coherent theory."[48]

B. Machine Learning Versus Traditional Computer Programming

In previous decades, machines operated according to rules that humans painstakingly programmed by hand, "writing code of exactly what [they] want[ed] the machine to do."[49] This method of computation powered all the wide array of computer applications through the twentieth century, but it could not automate the many tasks that humans do that cannot be practically reduced to sets of rules. One such task is facial recognition. Using the example of how he can easily recognize his mother's face, one AI researcher comments, "I . . . recognize it but I couldn't really write code to do it."[50] It is for this reason, according to Polanyi's paradox, that there are fundamental limits to how much knowledge humans can impart to machines.[51] More recently, however, machine learning has emerged as a revolutionary subfield of AI because it can circumvent that limitation.

In short, machine learning refers to a program's ability to "extract[ ] patterns from raw data."[52] "Deep learning," a type of machine learning, has powered much of the recent gains in machine learning research. Deep learning programs optimize accuracy and, over time, yield increasingly accurate results for a given task. That is, the machine has the "ability to keep improving its performance without humans having to explain exactly how to accomplish" a task.[53] Now, "machines learn on their own things that we don't know how to explain."[54] After being shown thousands or even millions of examples,[55] the machines learn patterns, correlations, and rules—

(2017) (discussing the AI AlphaGo's use of predictive technology to master the game of Go).

48 Jerry Kaplain, *AI's PR Problem*, MIT TECH. REV. (Mar. 3, 2017), https://www.technologyreview.com/s/603761/ais-pr-problem/.

49 *How AI Is Already Changing Business*, HBR IDEACAST (July 20, 2017), https://hbr.org/ideacast/2017/07/how-ai-is-already-changing-business [hereinafter *AI Changing Business*].

50 *Id.*

51 *See* Brynjolfsson & McAffee, *supra* note 37, at 6; *see also* David H. Autor, *Polanyi's Paradox and the Shape of Employment Growth* 8 (Nat'l Burea of Econ. Research, Working Paper No. 20485, 2014) ("[E]ngineers cannot program a computer to simulate a process that they (or the scientific community at large) do not explicitly understand. This constraint is more binding than one might initially surmise because there are many tasks that we understand tacitly and accomplish effortlessly for which we do not know the explicit 'rules' or procedures.").

52 IAN GOODFELLOW, YOSHUA BENGIO & AARON COURVILLE, DEEP LEARNING 2–3 (9th ed.2016).

53 Brynjolfsson & McAffee, *supra* note 37, at 4.

54 *AI Changing Business*, *supra* note 49.

55 *See* Yonghui Wu et al., *Google's Neural Machine Translation System: Bridging the Gap Between Human and Machine Translation*, ARXIV.ORG 1, 14 (2016), https://arxiv.org/pdf/1609.08144.pdf (describing the process of teaching a machine English to French translation using thirty-six million pairs of sentences).

sometimes the ones that humans use to accomplish the task but other times ones that humans cannot perceive,[56] or had not used previously.[57] Indeed, many times the programmer him- or herself cannot account for *how* the machine came to a particular result, even if the result is correct.[58] Tasks that were once impossible to automate are now on par with human experts, including not only facial recognition,[59] but also skin cancer detection[60] and some types of language translation.[61]

With many applications emerging, and far more on the horizon, it is inevitable that attorneys will do with machine learning what they have done before with all manner of devices, machines, and technical software: use it to win. Law firms are already incorporating machine learning software into other aspects of their business, like e-discovery,[62] while government regulators have begun to use machine learning to assist in investigating fraud and other white-collar crimes.[63] Prosecutors, specifically, may find several aspects of their work affected by machine learning, including justifying

---

56 *See, e.g.*, Heather Murphy, *Why Stanford Researchers Tried to Create a 'Gaydar' Machine*, N. Y. TIMES (Oct. 9, 2017), https://www.nytimes.com/2017/10/09/science/stanford-sexual-orientation-study.html (using photos of gay men and straight men, an AI was able to use aspects of the human face to predict a man's sexual orientation with up to ninety-one percent accuracy).

57 *Id.*

58 Andreas Holzinger et al., *What Do We Need to Build Explainable AI Systems for the Medical Domain?*, ARXIV.ORG (2017), https://arxiv.org/pdf/1712.09923.pdf ("However, the central problem of such models is that they are regarded as black-box models and even if we understand the underlying mathematical principles of such models they lack an explicit declarative knowledge representation, hence we have difficulty in generating the underlying explanatory structures.").

59 *See generally* Will Knight, *Paying with Your Face*, MIT TECH. REV. (Mar.–Apr. 2017), https://www.technologyreview.com/s/603494/10-breakthrough-technologies-2017-paying-with-your-face/ (detailing how researchers have shown their programs rival most humans in ability to recognize faces).

60 Andre Esteva et al., *Dermatologist-level Classification of Skin Cancer with Deep Neural Networks*, 542 NATURE 115, 115 (2017).

61 Wu et al., *supra* note 55, at 19.

62 *See* Avaneesh Marwaha, *Seven Benefits of Artificial Intelligence for Law Firms*, LAW TECH. TODAY (July 13, 2017), http://www.lawtechnologytoday.org/2017/07/seven-benefits-artificial-intelligence-law-firms/; Catalyst, *How AI and Machine-Learning Tools Lighten the eDiscovery Load*, ABOVE L. (May 17, 2017, 3:02 PM), https://abovethelaw.com/2017/05/how-ai-and-machine-learning-tools-lighten-the-ediscovery-load/; Julie Sobowale, *How Artificial Intelligence Is Transforming the Legal Profession*, A.B.A. J. (Apr. 2016), http://www.abajournal.com/magazine/article/how_artificial_intelligence_is_transforming_the_legal_profession (all outlining how AI is used to save time in discovery).

63 Gerard Hoberg & Craig Lewis, *Do Fraudulent Firms Produce Abnormal Disclosure?* 1–3 (Vand. Owen Graduate Sch. of Mgmt. Research Paper No. 2298302, 2015), (using a topic modeling technique that discovers clusters of text to predict whether a firm's SEC disclosure shows signs that the firm is committing fraud); Scott W. Bauguess, *The Hope and Limitations of Machine Learning in Market Risk Assessment*, SEC (Mar. 6, 2015), https://cfe.columbia.edu/files/seasieor/center-financial-engineering/presentations/MachineLearningSECRiskAssessment030615public.pdf (discussing how the SEC could produce a model to help detect illicit behavior).

searches[64] and determining which crimes to prosecute.[65] Though machine learning has not yet been widely used to produce evidence itself, the capability, accessibility, and incentives to do so already exist.

III. HOW MACHINE LEARNING WORKS

A machine learning program extracts useful patterns out of a large collection of data to perform a certain task.[66] To be clear, the learning itself is not the ultimate goal, but rather the means to achieve that goal.[67] "Learning," in this context, refers to an improvement in performance of the task over time.[68] Practicing attorneys can at least grasp the fundamentals of machine learning by becoming familiar with the tasks these programs can perform and the processes by which the machines "learn."

A. Tasks

Machines can learn to perform many tasks. The most common include classification (e.g., image or facial recognition), classification with missing inputs (e.g., recognizing an object or face from a corrupted or incomplete image), regression (e.g., predicting a numerical value given certain conditions), transcription (e.g., speech-to-text software), machine translation (e.g., translating from one natural language to another), structured output (e.g., image recognition in which the machine can describe the image in grammatical sentences), anomaly detection (e.g., credit card fraud detection), synthesis and sampling (i.e., the machine generates new examples similar to the examples it has learned), imputation of missing values (i.e., predicting certain data points given other data points), and denoising (i.e., match an inputted "corrupted" exampled to a "clean" example).[69]

Many of the emerging or possible evidentiary applications of machine learning fall into these general categories. For instance, risk assessment in parole hearings could be accomplished with a regression analysis.[70] Facial recognition could identify a defendant even with video or photographic

---

64 *See generally* Rich, *supra* note 11 (discussing machine learning and the Fourth Amendment).

65 *See* Ferguson, *supra* note 1, at 732 ("[T]he predictive prosecution model shifts the identification of problem areas from the street cops to the lawyers.").

66 Kaplan, *supra* note 48.

67 *See* GOODFELLOW ET AL., *supra* note 52, at 97 (explaining that once a specific task is defined, like walking, learning how to do the task is not the task itself, but gaining the means to perform the task).

68 *Id.* ("A computer program is said to learn from experience *E* with respect to some class of tasks *T* and performance measure *P*, if its performance at tasks in *T*, as measured by *P*, improves with experience *E*.").

69 *Id*. at 98–101.

70 Lacambra, *supra* note 3.

evidence in less than ideal circumstances.[71] Body recognition algorithms may achieve the same where no facial images are captured.[72] Anomaly detection can scan corporate filings or other behavior to assess evidence of wrongdoing.[73] It impossible to catalog all the ways in which machine learning may produce evidence, especially as the technology further evolves, but suffice to say these are only among the presently foreseeable.

B. Learning

To perform the task, the machine first must learn from examples, which are simply a collection of quantified features.[74] When the data are already numerical, quantification is straightforward. In other situations, how the data is quantified is not immediately obvious or can reflect programmer judgment. For instance, an image of a face is quantified on the basis of pixel values that a screen would use to display the image.[75] Once the data have been translated into numbers, the programmer must take some of the data whose properties are already known, referred to as "training data,"[76] and teach the machine the rules or associations that will be useful when the machine later analyzes new data whose properties are *not* already known. This process is referred to as "supervised learning."[77]

To echo the facial recognition example above,[78] a programmer at this stage will feed a set of pictures of her mother (which the programmer knows to be of her mother) into the machine. Critically, the programmer explicitly tells the machine to associate the images of that face with her mother, such as by labeling each image with the mother's name. At this point, the machine knows these images are of the mother not by any inference or computation, but because the programmer has told the machine explicitly. Then, the machine analyzes the pictures of the mother's face and, on its own, establishes other associations, correlations, or rules that will enable it to

---

71 See John Nawara, *Machine Learning: Face Recognition Technology Evidence in Criminal Trials*, 49 U. LOUISVILLE L. REV. 601, 608–09 (2011).

72 Chikahito Nakajima et al, *Full-body Person Recognition System*, 36 PATTERN RECOGNITION 1997, 1997 (2003) ("We describe a system that learns from examples to recognize person in images taken indoors.").

73 Hoberg & Lewis, *supra* note 63; Bauguess, *supra* note 63.

74 *See* GOODFELLOW ET AL., *supra* note 52, at 97 (defining example as "a collection of features that have been quantitatively measured from some object or event that we want the machine learning system to process")*; id.* at 103 (explaining how "supervised learning" in deep learning computers uses sets of data curated and labeled for the neural network to experience).

75 *Id.* at 97 ("[T]he features of an image are usually the values of the pixels in the image.").

76 *Id.* at 119.

77 *Id.* at 103. Note that these categories are not clearly defined and may blur at the edges.

78 *See* Nawara, *supra* note 71.

recognize the programmer's mother in new images it has not seen before. For example, the machine might establish rules about skin tone, distance of the eyes from one another, and height or width of the face.[79]

Once the machine has learned from the training data and deduced some set of rules, its performance is then tested and refined on a separate pool of testing data, called the "test set," the properties of which are also known.[80] The programmer then assesses the error rates of the machine's accuracy and makes adjustments. In the present example, our programmer would at this stage feed into the machine new images of her mother that the machine has not seen before and test how well it can identify the mother. When the machine has reached some level of accuracy that the researcher feels is satisfactory, it is used to analyze real world data. Ideally, the machine should be able to identify the mother in any image where she is present, including situations of various image quality, bright or dark lighting, different angles, or no matter the mother's hair style, presence or absence of makeup, differences in outfit, or other situations where her appearance is slightly different.

IV. ADMISSIBILITY UNDER FEDERAL RULE OF EVIDENCE 702

When machine learning output is used as substantive evidence in litigation in federal court, it most likely will be in the form expert testimony governed by Rule 702 and *Daubert*, though if or how it may be used in state courts depends on each state's rules of evidence.[81] Rule 702 governs the

---

79 *See* GOODFELLOW ET AL., *supra* note 52, at 8 (explaining that "deep learning" is a type of machine learning that is often used in facial recognition, but is used in other contexts, as well. In general, deep learning is a process of representing abstract concepts in terms of simpler concepts.); *see also id.* at 6 (indicating an abstract concept, like a human face, can be represented as a particular arrangement of simpler concepts, like lines, contours, and edges); *id.* at 8 (describing a typical deep learning algorithm would analyze an image first for a series of lines (a relatively simple analysis), then for a series of connected and contoured lines (a slightly more complex analysis building upon the first), and finally assess if the present arrangement of lines, contours, and edges matches the arrangement that the algorithm had learned corresponds to a face).

80 *Id.* at 106 (explaining how accuracy of the performance data is tested).

81 The Federal Rules of Evidence do not govern all the ways in which machine learning will likely infect litigation, such as investigations, bail determinations, and sentencing. *See* Robin A. Smith, *Opening the Lid on Criminal Sentencing Software*, DUKE TODAY (July 19, 2017), https://today.duke.edu/2017/07/opening-lid-criminal-sentencing-software (explaining the black-box software of sentencing as a mystery); Laura Smith-Spark, *Voice, Words May Provide Key Clues About James Foley's Killer*, CNN (Aug. 24, 2014), https://www.cnn.com/2014/08/22/world/europe/british-jihadi-hunt/index.html (last updated Aug. 24, 2014, 6;07 AM)) (discussing the use of voice identification software to identify a hooded ISIS militant); Eric Westervelt, *Did a Bail Reform Algorithm Contribute to This San Francisco Man's Murder?*, NPR (Aug. 18, 2017, 2:00 PM), https://www.npr.org/2017/08/18/543976003/did-a-bail-reform-algorithm-contribute-to-this-san-francisco-man-s-murder) (discussing a computer algorithm deciding the fate of a defendant). Without the schema of the Federal Rules of Evidence in place, it is safe to assume that those areas

admissibility of expert testimony in federal court.[82] For a qualified expert to testify, the proponent must show that the testimony will assist the trier of fact, that the opinion is based on sufficient facts or data, that the testimony is the product of reliable principles and methods, and that the principles and methods are reliably applied to the instant case.[83] When the judge determines the admissibility of expert testimony, she is only making "a preliminary assessment of whether the reasoning or methodology underlying the testimony is scientifically valid and of whether that reasoning or methodology properly can be applied to the facts in issue."[84] The focus is not on the conclusions that the methods generate.[85]

Machine learning output is likely admissible under both under *Daubert* and the text of Rule 702 itself. However, the exact manner in which the algorithm was created or the way it would be used at trial may, in some cases, render it inadmissible.

A. Daubert *Criteria*

In *Daubert v. Merrell Dow Pharmaceuticals*, the Supreme Court established a general framework for federal courts to assess whether expert testimony is the product of "reliable principles and methods" under Rule 702.[86] The Court lists four non-dispositive considerations, none of which categorically bar machine learning evidence. First, whether the theory or technique can be or has been tested; second, whether the theory or technique has been subject to peer-reviewed publication; third, the existence of error rates; and fourth, whether the theory or technique enjoys general acceptance in the

---

are the first where we would see machine learning at work, especially since today those areas are seeing *non*-machine learning algorithms grow in popularity and legal legitimacy. *See*, *e.g.*, State v. Loomis, 881 N.W.2d 749, 752–53 (Wis. 2016) (describing the use of risk assessment algorithms in the context of probation, parole, and sentencing); Malenchik v. State, 928 N.E.2d 564, 575 (Ind. 2010) (concluding that trial judges "are encouraged" to use risk assessment software to inform sentencing decisions).

82 Federal Rules of Evidence 702 reads:

> A witness who is qualified as an expert by knowledge, skill, experience, training, or education may testify in the form of an opinion or otherwise if:
>
> (a) the expert's scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue;
>
> (b) the testimony is based on sufficient facts or data;
>
> (c) the testimony is the product of reliable principles and methods; and
>
> (d) the expert has reliably applied the principles and methods to the facts of the case.

FED. R. EVID. 702.

83 *Id*.

84 Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 592–93 (1993).

85 *Id.* at 595 (stating the focus is solely on principles and methodology).

86 *Id*.; *see also* FED. R. EVID. 702(c).

field or scientific community.[87]

Machine learning easily satisfies three of the four *Daubert* factors without extensive discussion. Machine learning evidence would certainly meet the testability consideration, since these processes produce results that can be shown to be false, sometimes in spectacular ways. For instance, in 2015, Google's object recognition system falsely identified two African Americans as gorillas, quickly prompting outcry and a correction to the algorithm.[88] Machine learning also satisfies *Daubert*'s peer review consideration, since the peer reviewed literature on it has proliferated in recent years, with some of its scientific principles dating back to the mid-twentieth century.[89] And, machine learning enjoys general acceptance in the field or scientific community, and practitioners are applying the technology in myriad disciplines.[90]

Daubert's requirement that the science have either known or potential error rates,[91] however, presents a more complicated analysis. Machine learning algorithms indeed have calculable error rates, though the relevance of these error rates to the particular situation is oftentimes questionable.

Machine learning algorithms usually have two important error rates. The first is its test set error rate with respect to training data, which are the examples whose properties are already known to the researcher and which are the basis for the algorithm's improved performance over time.[92] Eventually, a second error rate captures the algorithm's performance when it is unleashed upon real-world examples with unknown properties.[93] Both error rates typically appear as a singular number that masks other important statistics, like whether the algorithm is more likely to give false positives or false negatives, an important detail that should be revealed at a *Daubert* hearing or on cross examination.

Subjective programmer judgments can inform the error rate, such as whether or not to give partial credit for a partial success,[94] though in some contexts it is difficult to assess what should be considered a success or failure

87 *Daubert*, 509 U.S. at 593–94.

88 Tom Simonite, *When It Comes to Gorillas, Google Photos Remains Blind*, WIRED (Jan. 11, 2018, 7:00 AM), https://www.wired.com/story/when-it-comes-to-gorillas-google-photos-remains-blind/ ("In 2015, a black software developer embarrassed Google by tweeting that the company's Photos service had labeled photos of him with a black friend as 'gorillas.'").

89 *See* GOODFELLOW ET AL., *supra* note 52, at 12. ("[D]eep learning dates back to the 1940s.").

90 *Id.* at 98–101.

91 *Daubert*, 509 U.S. at 594 (stating in the case of a scientific technique the court should consider the potential rate of error).

92 *See* GOODFELLOW ET AL., *supra* note 52, at 102 (referring to the error rate value as "the expected 0-1 loss").

93 *Id.*

94 *Id.*

in the first place. For example, in a lip-reading algorithm, is an inelegant but understandable translation a success or a failure? And if it is only a partial success, how partial is it? The answer, which will inform the error rates, is ultimately a human judgment, and there may be no consistency from one programmer to another. For purely binary outcomes, like the task of identifying a defendant, no such thing as partial success would exist, because the individual the algorithm is identifying in a video, photo, or recording either is the defendant or is not.

Additionally, a machine's overall stated error rate may mask a higher rate of error when it draws conclusions about a defendant who does not share characteristics with the initial training data. For instance, an error rate for a machine that has been trained on racially diverse data may be less reliable for a single racial category than others. In one facial recognition application, "the software is right 99 percent of the time" but only "[w]hen the person in the photo is a white man."[95] "But the darker the skin, the more errors arise—up to nearly 35 percent for images of darker skinned women."[96] Yet, oftentimes today's machines are *not* trained on racially diverse data, which presents other problems for how to generalize its conclusions. For instance, one recent facial recognition system reported 97.35% accuracy but on a dataset that turned out to be 77.5% male and 83.5% white.[97] Its error rates were never broken down by race or gender.[98]

Aurally, too, machines struggle with accents that are not standard American or British. Speech recognition algorithms may vary in their accuracy when dealing with accents from various regions. Scottish was the most difficult for one speech recognition algorithm to understand, followed closely by American southerners from Georgia.[99] Nor are these variables entirely independent. Sometimes the accuracy of a speech recognition algorithm is highly correlated with race, gender, or age: "higher-pitched voices are more difficult for speech-recognition systems" which makes them

---

95 Steve Lohr, *Facial Recognition Is Accurate, If You're a White Guy*, N.Y. TIMES (Feb. 9, 2018), https://www.nytimes.com/2018/02/09/technology/facial-recognition-race-artificial-intelligence.html?smid=fb-nytimes&smtyp=cur.

96 *Id*.

97 Joy Buolamwini & Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, 81 PROC. MACHINE LEARNING RES. 1, 3 (2018) (citing Hu Han & Anil K. Jain, *Age, Gender and Race Estimation from Unconstrained Face Images*, MSU TECH. REP. 1, 2 (2014)).

98 *Id.* at 3 (citing Yaniv Taigman et al., *Deepface: Closing the Gap to Human-level Performance in Face Verification*, 2014 IEEE CONF. COMPUTER VISION & PATTERN RECOGNITION 1701, 1701).

99 Johnson, *In the World of Voice-recognition, Not All Accents Are Equal*, ECONOMIST (Feb. 15, 2018), https://www.economist.com/news/books-and-arts/21737017-you-can-train-your-gadgets-understand-what-youre-saying-world ("The automatic captioning did worst with the Scottish speakers, transcribing more than half of the words incorrectly, followed closely by American southerners (from Georgia).").

less accurate overall for women and especially children.[100] Multiple popular speech recognition algorithms had similar trouble with black and mixed-race speakers.[101]

Thus, the mosaic of different possible error rates presents a more complicated picture than a single, impressively low error rate may reflect. For this reason, machine learning evidence is particularly susceptible to violating Rule 702(d)'s requirement that the evidence be "reliably applied the principles and methods to the facts of the case."[102] If an algorithm has an impressive rate of error with respect to data that bears little resemblance to the instant defendant, then its conclusions are not being reliably applied to the facts of the case.[103]

B. Problems of Data

Rule 702 requires that the proffered evidence be based on sufficient facts or data and be the product of reliable principles and methods.[104] This section suggests several inquiries of data collection and use that may affect the admissibility of machine learning output under 702(b) and 702(c).

1. How Large Was the Training Dataset?

Sample size is an initial inquiry that is by no means unknown to lawyers challenging scientific evidence.[105] Machine learning algorithms require very large datasets to extract useful patterns and make accurate assessments, and more complicated tasks require more examples to fine tune their accuracy. For instance, text recognition (a relatively simple task) may require only a few thousand examples, whereas language translation (an extremely complex task) requires tens of millions of examples.[106] The party seeking to admit the evidence would want assurances that the training data is sufficiently large for the given task, whereas the party seeking to exclude the evidence would want

---

100 *Id*. ("It also did worse with women: higher-pitched voices are more difficult for speech-recognition systems, one reason they tend to struggle with children.").

101 *Id*. ("In a follow-up experiment, Ms. Tatman used both YouTube and Bing Speech, made by Microsoft, to test only American accents. Both found black and mixed-race speakers harder to comprehend than white ones.").

102 FED. R. EVID. 702(d).

103 Gen. Elec. Co. v. Joiner, 522 U.S. 136, 146 (1997) ("A court may conclude that there is simply too great an analytical gap between the data and the [expert] opinion proffered.").

104 FED. R. EVID. 702(b)–(c).

105 *See* FED. JUD. CTR., REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 246 (3d ed. 2011) (pondering the question of how large a sample size should be when lawyers are making scientific inquiries).

106 *See* Wu et al., *supra* note 55 (teaching a machine English to French translation using thirty-six million pairs of sentences).

to inquire as to how many examples the algorithm has learned and if that number is in keeping with what is generally accepted for the task.

2. Were the Training Data Gathered or Generated in Ways that Produced a Biased Sample?

Not only must the dataset be large, but it also must have some baseline quality to make useful predictions. The quality of the data, and the extent to which it may be biased in a particular way, can be probed with various inquiries. Where did the data come from? Did the researcher him-or-herself gather the data according to accepted methods? If the researcher instead received the data from a third party, can he or she vouch for its quality in any specific way? In the case of open source methods[107] or crowdsourced[108] data, which are common in the machine learning field, is such verification even possible?

Courts are already familiar with challenges to data collection methods, and evaluating whether they produced a biased sample that would reduce the data's relevancy to the present issue. In the case of machine learning, the representativeness of the dataset to the given defendant's jurisdiction, crime, or any other feature is crucial for drawing appropriate conclusions from the machine's output. This inquiry has obvious implications for a potential Equal Protection challenge,[109] but even assuming there are no cognizable constitutional issues with the data, the data simply may not be relevant to a given defendant for any number of reasons.

The Wisconsin Supreme Court in *State v. Loomis*[110] demonstrates how these bias and relevancy concerns are already manifesting in algorithmic output based on data. In challenging the State's use of Correctional Offender Management Profiling for Alternative Sanctions ("COMPAS") to determine his sentence, Loomis's expert testified that sentencing courts have little assurance that the data COMPAS uses are unbiased, or were even

107 "Open source software is software with source code that anyone can inspect, modify, and enhance." *What Is Open Source?*, OPEN SOURCE, https://opensource.com/resources/what-open-source (last visited Oct. 24, 2018).

108 "Crowdsourcing is a type of participative online activity in which an individual, organization, or company with enough means proposes to a group of individuals of varying knowledge, heterogeneity, and number, via a flexible open call, the voluntary undertaking of a task." Enrique Estellés Arolas & Fernando González Ladrón-de-Guevara, *Towards an Integrating Crowdsourcing Definition*, 38 J. INFO. SCI. 1, 11 (2011).

109 The extent to which machine learning evidence might be sufficiently biased in a way that is adverse to minority groups to mount a cognizable Fourteenth Amendment challenge is outside the scope of this Comment, but it is a critical question ripe for further research.

110 881 N.W.2d 749, 754 (Wis. 2016) (noting that the risk-need assessment tool name COMPAS does not predict the specific likelihood that an individual offender will reoffend).

relevant to Loomis:

> The Court does not know how the COMPAS compares that individual's history with the population that it's comparing them with. The Court doesn't even know whether that population is a Wisconsin population, a New York population, a California population. . . . There's all kinds of information that the court doesn't have, and what we're doing is we're misinforming the court when we put these graphs in front of them and let them use it for sentence.[111]

Similar questions would be appropriate when probing a machine learning dataset's relevancy. If a machine learning algorithm is generating inculpatory evidence for a Wisconsin defendant, should the data only come from the Wisconsin population, or the Midwest region, or can nationwide data suffice? Geography would not be the only consideration, as other factors could bias the data. The answers to these questions are intensely fact-specific and would depend on what the proponent of the machine learning evidence is trying to prove.

Moreover, even if the prosecution relies on official statistics gathered by government agencies, these datasets are not inherently high quality. Today, the accuracy of government databases is often accepted "as an article of faith, with courts according them a presumption of reliability."[112] While data-driven governance is often a laudable goal, "[t]oday, the prevailing zeitgeist of governments is one of database expansion, not quality control or accountability, and a blasé acceptance of data error and its negative consequences for individuals."[113] Some important figures have taken note. In *Herring v. United States*, Justice Ginsburg recognized in her dissent that "[t]he risk of error stemming from these databases is not slim," noting issues with National Crime Information Center, terror watch lists, and public employment databases.[114]

Professors Logan and Ferguson note the series of challenges and pitfalls that accompany government database creation. When data are first gathered or generated, basic human error in collection or interpretation is common.[115] Sometimes data are collected and uploaded without legal authorization or counter to what was initially ordered.[116] Once errors are

---

111 *Id.* at 756–57 (quoting the testimony of Loomis's expert).

112 Wayne A. Logan & Andrew Guthrie Ferguson, *Policing Criminal Justice Data*, 101 MINN. L. REV. 541, 543–44 (2016).

113 *Id.* at 543.

114 555 U.S. 135, 155 (2009) (Ginsburg, J., dissenting).

115 Logan & Ferguson, *supra* note 112, at 559 ("At the point of [data] collection, accuracy can be impaired by basic human error.").

116 *Id.* (noting that states often upload DNA profiles not authorized by the law and DNA information that should be destroyed is often retained).

made, they are difficult to discover and difficult to correct.[117] If the error is corrected in one database, it is not guaranteed that the correction will filter to the myriad of other databases that had, in the past, copied from the initial database.[118] Of course, in a federalist system with hyperlocalist police power, uniform data collection, management, correction, and dissemination would be as difficult to implement as it would be helpful.[119]

3. Was the Data Manipulated? If So, How, and Does that Matter?

When a dataset is not large enough, programmers have several techniques for manipulating it to artificially create a larger training set. For example, the algorithm may take many random samples from the original dataset to create many other, smaller datasets.[120] The programmer may also intentionally distort the examples, such as by warping images or adding random noise.[121] The forms of manipulation are largely influenced by subjective programmer judgment and norms in the field.

4. How Was the Data Tagged and Labeled?

Moreover, even if a large dataset is collected or generated using standard techniques, it must be labeled and organized properly, which, for datasets with millions of examples, is a menial but crucial task. Machine learning programs only "learn" what they are "taught" from the data, and it is the programmers who make judgments about what the data show by the way that they are labeled. Indeed, researchers can intentionally teach the algorithm nonsense simply by labeling.[122] In that way, who labeled the data and how—and the extent to which the labeling was done properly—are

---

117 *Id.* at 586 ("Ex ante detection of database error, as Professor Kenneth Karst noted fifty years ago, 'depends on the subject's access to his own file and his awareness of the need to inspect it. Even when a record is freely accessible to its subject, there is no assurance that the subject will know of its existence or its contents.'" (quoting Kenneth L. Karst, *"The Files": Legal Controls over the Accuracy and Accessibility of Stored Personal Data*, 31 LAW & CONTEMP. PROBS. 342, 358 (1966)).

118 *Id.* at 588 ("Data is often shared, replicated, backed up and stored in many different databases at once. Even if a data error is corrected, this does not guarantee that other shared datasets will reflect the change.").

119 *Id.* at 596–611 (suggesting legislation, regulation, and best practices to coordinate data at the federal, state, and local levels).

120 GOODFELLOW ET AL., *supra* note 52, at 120 (discussing the most common method, the k-fold cross-validation procedure).

121 *See* Ritchie Ng, *Machine Learning Photo OCR*, RITCHIENG.COM http://www.ritchieng.com/machine-learning-photo-ocr/ (last updated Oct. 13, 2018) (noting the possibility of distorting examples through "warping the image").

122 *See* 3Blue1Brown, *Gradient Descent, How Neural Networks Learn | 2, Deep Learning, Chapter 2*, YOUTUBE (OCT. 16, 2017), at 18:10, https://www.youtube.com/watch?v=IHZwWFHWa-w (teaching an algorithm that an image of Isaac Newton is an image of a cow).

critical inquiries.

While the researcher may do the labeling herself, it is unlikely that she is labeling millions of examples by hand. Often researchers use open datasets already created for public use, but the researcher may have no idea how that data set was created and labeled.[123] Strangers sitting at home may do it for nominal payment via Amazon Mechanical Turk.[124] And, if one machine learning algorithm can label data,[125] other machine learning algorithms can then use that labeled data to learn other tasks, which can clearly have the advantage of labeling quickly but would only further compound the potential risks by adding one machine learning process on top of another.[126]

C. Problems in the Source Code

An examination of software's source code may also bring to light details that affect the admissibility of the evidence under Rule 702. If the programming itself contains errors, then it is possible that the program's conclusions are not the "product of reliable principles and methods."[127] Broadly speaking, "source code" is a combination of words and mathematical symbols that have a particular meaning in a programming language.[128] Unlike "machine code," which is a binary collection of 1's and

---

123 Hector Garcia-Molina et al., *Challenges in Data Crowdsourcing*, 28 IEEE TRANSACTIONS KNOWLEDGE & DATA ENGINEERING 901, 905–07 (discussing problems with crowdsourced data).

124 *See* Ng, *supra* note 121 ("Hire people on the web to label data (amazon mechanical turk).").

125 Tom Simonite, *Google's Brain-Inspired Software Describes What It Sees in Complex Images*, MIT TECH. REV. (Nov. 18, 2014), https://www.technologyreview.com/s/532666/googles-brain-inspired-software-describes-what-it-sees-in-complex-images/ ("Researchers at Google have created software that can use complete sentences to accurately describe scenes shown in photos . . . .").

126 Linking machine learning applications in this way is increasingly common. One University of California, Berkeley researcher has developed a dual machine learning system in which one algorithm identifies the species of bird in a photograph, while a second algorithm analyzes the decision-making of the first and creates, in sentence format, explanations of how the first algorithm made its species determination. *See For Artificial Intelligence to Thrive, It Must Explain Itself*, ECONOMIST (Feb. 15, 2018), https://www.economist.com/news/science-and-technology/21737018-if-it-cannot-who-will-trust-it-artificial-intelligence-thrive-it-must (discussing the pros and cons of "deep learning" in artificial intelligence). Google's AutoML project is actively researching machine learning algorithms that can themselves write new machine learning algorithms. *See* Cade Metz, *Building A.I. that Can Build A.I.*, N.Y. TIMES (Nov. 5, 2017), https://www.nytimes.com/2017/11/05/technology/machine-learning-artificial-intelligence-ai.html?_r=0 (discussing Google's search for artificial intelligence that can effectively build other A.I.-reliant mechanisms in the absence of human A.I. experts).

127 FED. R. EVID. 702(b).

128 Edward J. Imwinkelried, *Computer Source Code: A Source of the Growing Controversy over the Reliability of Automated Forensic Techniques*, 66 DEPAUL L. REV. 97, 104 (2016) ("The source code itself is a combination of words and mathematical symbols that have a particular meaning in the selected language.").

0's, the source code is human readable,[129] and is likely to be intelligible to a defense expert.[130] Source code dictates which tasks a computer program performs, how the program performs the tasks, and the sequence in which the program performs the tasks.[131] The source code can provide uninhibited access to the exact ways the programmer decided the machine will operate and is much more informative than simply observing what goes in and what comes out.[132]

Crucially, the source code can reveal simple errors or faulty assumptions in the program's creation. In a given program, millions of lines of code—often pieced together from innumerable sources and developers—give rise to simple accidents in transcription, mistakes in conditional programming, software rot,[133] or faulty updates to legacy code.[134] When one programmer designs the initial version of a program, it may be difficult for subsequent programmers in later versions to work around or adapt to the personal style and conventions of the first.[135] Studies demonstrate that, as a result, error rates of one percent in code are common, which can correspond to tens of thousands of errors in a single program.[136]

Moreover, sometimes the software itself contains no errors in the programming, but, because of human errors in communication or misunderstanding, the program does not accomplish the task that was ultimately sought.[137] When the device uses several different scientific disciplines—like, for example, the way a breathalyzer must incorporate knowledge from programming, chemistry, and biology—differences in

---

129 *Id.* at 105.

130 *Id.*

131 *Id.* at 103.

132 Chessman, *supra* note 19, at 182 ("While some information can be gleaned from viewing the program in action, this information is highly limited and may omit crucial details that relate to the reliability and accuracy of the program's output.").

133 *Id.* at 190 ("'Software rot' [happens] where the quality, functionality, and usefulness of a program actually degrade over time. . . . [It] occurs for a variety of reasons. At the most basic level, each software update creates new interactions between different portions of the source code, which may also entail unforeseen interactions and unforeseen consequences.").

134 *Id.* at 186–92.

135 *Id.* at 186 n.32 ("Subjective expressiveness is so pronounced that computer code is actually expressively distinguishable—it is possible 'to recognize the author of a given program based on programming style' in the same way one might identify Nietzsche by his obscurity or Hemingway by his verbosity." (quoting Jane Huffman Hayes & Jeff Offutt, *Recognizing Authors: An Examination of the Consistent Programmer Hypothesis*, 20 J. SOFTWARE TESTING VERIFICATION & RELIABILITY 329 (2010))).

136 *Id.* at 186–87.

137 *Id.* at 188. ("Even a programmer who makes no technical coding errors will produce inaccurate software if the programmer misunderstands the nature or requirements of the job. For example, a human programmer may misunderstand the program requirements because of miscommunication, misunderstanding, or accidental omission of important details during instruction.").

understanding can give rise to methodological errors that do not come to light until even after product launch.[138] In that case, the programming itself could be flawless, yet the machine would still be unreliable.

These issues have come to light in only the few cases where state supreme courts ordered comprehensive inspection into the reliability of certain devices. In a Minnesota inquiry into the Intoxilyzer 5000EN, a breathalyzer device, several reliability issues were uncovered with an examination of the source code. Specifically, it was discovered that the device "has a margin of error, that radio frequencies from cell phones can disturb the accuracy of the test, and that the test may erroneously produce a deficient sample."[139] Similarly, in New Jersey, a Special Master was appointed to evaluate the source code of the State's widely used breathalyzer device, the Alcotest 7110 MKIII–C.[140] While the device was ultimately found to be reliable in most cases, the Special Master uncovered several problems with how the device functioned in certain situations, such as when testing the blood alcohol content of women over sixty[141] in addition to other issues, like a need for a corrective multiplier for some temperature readings.[142] Importantly, none of these errors or considerations would have come to the attention of the court without examination of the source code.

Even while these issues present themselves in the context of traditional, non-machine learning software, there is little reason to think that machine learning program development is immune from human misunderstanding, slips of the finger in transcription, faulty assumptions, or biases. It is true that machine learning algorithms work differently than programs of the past, with bigger sets of data, more processing power, and a different methodology. However, they are still created according to the ways that all software is created: as a product of human decision making, with lines of code running in conjunction with other software, and on hardware that degrades with time.

D. Trade Secret Protections

As a result of the considerations above, lawyers will have a profound interest in examining the underlying data and source code of machine learning software for such errors—and yet, standing in their way will be trade secret protections and reluctance of courts to compel discovery into these

---

138 *See id.* at 188 n.48 (explaining how programmers of a breathalyzer used an incorrect conversion factor that was not discovered until examination of the source code).

139 *In re* Source Code Evidentiary Hearings in Implied Consent Matters, 816 N.W.2d 525, 545 (Minn. 2012).

140 State v. Chun, 943 A.2d 114, 120 (N.J. 2008).

141 *Id.* at 140.

142 *Id*. at 145.

possible defenses. Tech firms are particularly concerned with protecting trade secrets in machine learning because the field is still in its infancy, meaning that established players have less advantage over competitive startups than in other areas they typically dominate, like search in the case of Google or social media in the case of Facebook.[143] Today, in non-machine learning software, parties cannot observe the critical details of how the program was constructed because of its proprietary nature, and the programming firms themselves are often reluctant to reveal the source code or data that form the basis of their business success.[144]

A trade secret is nonpublic information that is the subject of reasonable efforts to maintain its secrecy and that confers a business advantage over competitors who lack that information.[145] Both data and source code have consistently been held to be trade secrets, and thus courts have often been reluctant to compel discovery into either, even for defendants in criminal actions who could use the information to mount a meaningful defense.

Defendants and third-party developers are increasingly disputing the discoverability and trade secret protections with respect to discovery of non-machine learning software, yet rarely is the source code turned over for inspection.[146] For instance, the two technologies that have so far experienced the most litigation over discovery of source code are infrared breath testing devices (i.e., breathalyzers) and DNA probabilistic genotyping, mostly surrounding the popular software TrueAllele.[147] In the breathalyzer cases, "the clear majority of courts rejected defendants' requests that a defense expert be granted access to the program's source code."[148] Likewise, in the TrueAllele cases, "although the issue has been litigated in at least seven states, no state court has ordered discovery of the TrueAllele source code" due to trade secret protections.[149]

---

143 *See Battle of the Brains*, ECONOMIST, Dec. 9, 2017, at 61, 62 (discussing how tech giants are investing large sums to develop their AI capabilities).

144 Chartes Tait Graves & Brian D. Range, *Identification of Trade Secrets Claims in Litigation: Solutions for a Ubiquitous Dispute*, 5 NW. J. TECH & INTELL. PROP. 68, 85–86 (2006) (describing the typicality of plaintiff corporations resisting to specifically disclose trade secret details in litigation, including research and development details and business strategies).

145 *See*, *e.g.*, 18 U.S.C. § 1839 (2012) (defining "trade secret"); Metallurgical Indus. Inc. v. Fourtek, Inc., 790 F.2d 1195, 1199–1203 (5th Cir. 1986) (explaining the requirements of a protected trade secret).

146 Imwinkelried, *supra* note 128, at 100.

147 *Id.*

148 *Id.* (citing State v. Underdahl, 749 N.W.2d 117, 120–21 (Minn. Ct. App. 2008), *aff'd in part, rev'd in part*, 767 N.W.2d 677 (Minn. 2009) (affirming the district court's denial of production of computer code); People v. Cialino, 831 N.Y.S.2d 680, 681–82 (N.Y. Crim. Ct. 2007); State v. Burnell, No. MV06479034S, 2007 WL 241230, at *2 (Conn. Super. Ct. Jan. 18, 2007); State v. Walters, No. DBDMV050340997S, 2006 WL 785393, at *1 (Conn. Super. Ct. Feb. 15, 2006); Moe v. State, 944 So. 2d 1096, 1097 (Fla. Dist. Ct. App. 2006)).

149 Imwinkelried, *supra* note 128 at 111.

Undoubtedly, the state has a legitimate interest in protecting trade secrets not only for developers' economic protection, but also to ensure society reaps the benefit of continued innovation. Trade secrets are protected in federal and state statutes, as well as incorporated into Rule 501's evolving common law of privileges.[150] And yet, under statute and at common law, it is well-settled that the trade secret privilege is a conditional or qualified one.[151] The trade secrets protections of every state include some form of an "injustice exception" that allows for discovery. "While the precise wording varies from state to state, the injustice exceptions substantially suggest that trade secret privilege from discovery exists only 'if the allowance of the privilege will not tend to conceal fraud or otherwise work injustice.'"[152]

Thus, courts have a number of tools at their disposal to not only allow source code discovery in the first place (permission by statute or Rule 501), but also to protect the legitimate economic interests of developers. Once discovery is compelled, courts have several safeguards to protect developers' business interest: courts can conduct in camera review, issue protective orders, seal records, threaten sanctions for improper disclosure, or require the parties to mutually agree on a third-party to review the source code.[153] Unfortunately, courts rarely use these tools and instead typically deny discovery altogether.[154]

The California Court of Appeals, reasoning in *People v. Superior Court (Chubbs)*, typifies how courts often hold that discovery of the source code itself requires meeting a high burden that the source code will assist the defense where no other unprotected information will. In that case, Martell Chubbs was charged with murder on the basis of a DNA result that would on average match randomly 1 in 10,000 times.[155] At trial, however, the prosecution put forward a different analysis that put the match as randomly occurring on

---

150 *Id*. at 125 ("Although Congress balked at enacting the draft rule, many states have done so; regardless, the federal courts have recognized the privilege by common-law process under Federal Rule 501.").

151 *Id*. at 126.

152 Chessman, *supra* note 19, at 212 (quoting JEROME G. SNIDER ET AL., CORPORATE PRIVILEGES AND CONFIDENTIAL INFORMATION § 8.02[1] (2011)).

153 *See id.* at 213. Many other forms of protection are also possible. In civil cases, courts have issue the following protective orders, inter alia: The opposing party's experts could examine the trade secret information only in a secure room; to gain access to the secure room, the experts had to identify themselves by iris and palm-print scans; during their examination of the information, the experts had to use paper bearing tags emitting radio waves to determine how many pages of notes the experts had used; counsel and the experts had to sign declarations that they would access the data only for use in the present litigation; and the trial courtroom would be closed to the public during any testimony discussing the trade secret information. *Id.*

154 Imwinkelried, *supra* note 128, at 126–27.

155 People v. Superior Court, No. B258569, slip op. at 3 (Cal. Ct. App. Jan. 9, 2015).

average 1 in 1.62 quintillion times.[156] Chubbs sought the source code of the subsequent program to account for the discrepancy and examine the assumptions built into the software.

The Court of Appeals held that source code is a trade secret and that it could be discoverable only by making "a prima facie, particularized showing" that the source code would be relevant and necessary to a defense.[157] The court concluded that Chubbs had not met that burden. The court reasoned that Chubbs had already received extensive information regarding the program's methodology and underlying assumptions from materials other than the source code.[158] Unfortunately, the appellate court did not explain how Chubbs could make the particularized showing it demanded or what would constitute sufficient particularity to overcome the trade secret protection.[159]

V. ADMISSIBILITY UNDER THE CONSTITUTION

Several constitutional provisions may be implicated by machine learning identification in criminal prosecutions. Defendants may cite the Fifth Amendment's Due Process Clause[160] or the Sixth Amendment's Confrontation Clause.[161] Some will likely provide little protection to

---

156 *Id.* at 3–4.

157 *Id.* at 10.

158 *Id.* at 21. This argument, that the defendant's access to other records, checklists for operation, and use manuals is sufficient to challenge the evidentiary weight of the device, is a common refrain in the courts. However, Professor Imwinkelreid argues these types of records are not nearly as informative as the source code. Commenting on similar reasoning of the Court in People v. Robinson, 860 N.Y.S.2d 159 (N.Y. App. Div. 2008), Professor Imwinkelreid argues:

> Those records do not contain the same information that an examination of the software's source code would yield. The analyst's checklist might minimize the risk of human error in conducting a test at a specific time and place, but the checklist provides no insight into any inherent defects in the program logic. Likewise, maintenance records could prove that for a certain period after a maintenance the device was operating as intended; but again, even if the device was operating as intended, there might be a defect buried in the source code. In sum, the discoverability of those documents does not undercut the case for discovery of the source code.

Imwinkelried, *supra* note 128, at 120.

159 Chessman, *supra* note 19, at 199 ("The appellate court did not explain how Mr. Chubbs could make the particularized showing it demanded without access to the source code, nor did it identify what showings would constitute sufficient particularity.").

160 The Fifth Amendment's Self-Incrimination Clause, and the Supreme Court's relevant jurisprudence, almost surely would allow prosecutors to require suspects to have recordings of their voice, images of their face, or other identifiers to be collected and fed into a machine learning algorithm, and hence this issue is not extensively discussed in this Comment. In short, because *Schmerber v. California* holds that only "testimony" may not be compelled under the Fifth Amendment, the state may compel physical evidence and identifiers that could be fed into the algorithm. 384 U.S. 757, 764 (1966).

161 There is much to explore with machine learning in the Fourth Amendment context that is mostly

defendants who wish to exclude inculpatory machine learning evidence, such as the Fifth Amendment. The Sixth Amendment, meanwhile, will almost surely require the evidence to admitted in the form of expert testimony but will not bar it entirely.

A. Due Process Under the Fifth Amendment

Machine learning output is often inexplicable, and experts sometimes cannot explain how the machine came to a particular conclusion.[162] On this basis, defendants may argue that such "guilt by black box" violates the Fifth Amendment's Due Process Clause[163] because, arguendo, "it offends some principle of justice so rooted in the traditions and conscience of our people as to be ranked as fundamental."[164] This "fundamental principle" may be that the inculpatory evidence must have some kind of discernible logic, explanation, ability to be examined or challenged. However, defendants making this argument will have little chance of success, at least as free-standing due process precedent currently exists.

Two background norms govern the Supreme Court's consideration of free-standing due process. First, where all the specific guarantees of the Bill of Rights have been observed and a guilty verdict has been reached, the Court typically finds that the defendant has thus enjoyed "all the process that is due."[165] "Where a particular Amendment 'provides an explicit textual

---

outside the scope of this Comment. *See generally* Melanie Reid, *Rethinking the Fourth Amendment in the Age of Supercomputers, Artificial Intelligence, and Robots*, 119 W. VA. L. REV. 863 (2017) (offering a new perspective on Fourth Amendment protections in the age of machine learning). However, there may be conceivable instances where a defendant would try to cite the Fourth Amendment as a bar to inculpatory evidence derived from machine learning processes. For example, to echo the facts of *Maryland v. King*, an individual may be arrested for one crime and have his photo taken, voice recorded, or cheek swabbed to gather data that would be fed into a machine learning algorithm, which could then identify and tie the arrestee to past unsolved crimes. 569 U.S. 435, 441 (2013). The Supreme Court held in *King* that a very similar situation was undoubtedly a search, and indeed one performed without individualized suspicion. *Id.* at 446, 448. But the Court ultimately held that it did not violate the Fourth Amendment because the state's interest in identifying perpetrators of past crimes outweighed the relatively non-invasive nature of a cheek swab for an arrestee. *Id.* at 453. Indeed, the mere photographic identification that would likely be employed in at least some machine learning analysis is undoubtedly even less invasive than that. "[W]e have never held that merely taking a person's photograph invades any recognized 'expectation of privacy.'" *Id.* at 477 (Scalia, J., dissenting) (quoting Katz v. United States, 389 U.S. 347 (1967)). Thus, defendants would likely find little help in the Court's Fourth Amendment jurisprudence to exclude machine learning evidence procured and used in this way.

162 For a more detailed look at explainability problems in machine learning, see *infra* Section VI.C.

163 "No person shall be . . .deprived of life, liberty, or property, without due process of law." U.S. CONST. amend. V.

164 Medina v. California, 505 U.S. 437, 445 (1992).

165 Daniel J. Steinbock, *Data Matching, Data Mining, and Due Process*, 40 GA. L. REV. 1, 23 (2005) ("Although the Due Process Clause provides a general baseline of fundamental fairness in the

source of constitutional protection' against a particular sort of government behavior, 'that Amendment, not the more generalized notion of 'substantive due process,' must be the guide for analyzing these claims.'"[166] Second, beyond enumerated protections in the Bill of Rights, whatever remaining protections are afforded by the free-standing Due Process Clause are "to be construed narrowly," and the Court has consistently declined to expand its scope, especially when doing so would interfere with the law enforcement powers of the states.[167]

No specific guarantee of the Bill of Rights regulates the admissibility of evidence, but the Due Process Clause does in very limited circumstances.[168] The Court currently recognizes only two forms of "bad evidence" that "invalidate the defendant's conviction on due process grounds."[169] One is "government-induced perjury"[170] and the other is "identification testimony from a suggestive lineup."[171] "Admission of such evidence, in the Court's view, is fundamentally unfair and violates due process."[172] This analysis, however, only demonstrates that free standing due process is the proper inquiry to evaluate evidence; once the Court is operating within that doctrine, defense counsel would face severe headwinds in trying to establish a free standing due process right to exclude machine learning evidence.

To do so, the defendant would have to show that machine learning evidence "offends some principle of justice so rooted in the traditions and conscience of our people as to be ranked as fundamental."[173] The inquiry of

---

criminal process, the Supreme Court has repeatedly held that where a more specific provision of the Bill of Rights applies, that provision constitutes all the process that is due."); *see also* Jerold H. Israel, *Free-Standing Due Process and Criminal Procedure: The Supreme Court's Search for Interpretive Guidelines*, 45 ST. LOUIS U. L.J. 303, 399 (2001) ("The range of regulation imposed under free-standing due process could also be restricted by giving a preemptive impact to the incorporated specific guarantees of the Bill of Rights.").

166 Albright v. Oliver, 510 U.S. 266, 273 (1994) (citing Graham v. Connor, 490 U.S. 386, 395 (1989)).

167 Israel, *supra* note 165, at 387 ("The Court's decisions in the post-incorporation era have, indeed, considered several major reformulations of due process doctrine as applied to criminal procedure, with some accepted and some rejected. The most important of these involved: (1) characterizing free-standing due process as a disfavored concept to be construed narrowly . . . ."). *But see id.* at 389–97 (questioning how truly "limited" free standing due process is by listing the "extraordinary range" of approximately fifty-one different due process protections the Court has found throughout the typical criminal adjudication timeline).

168 Alex Stein, *Constitutional Evidence Law*, 61 VAND. L. REV. 65, 86 (2008) ("The Due Process Clause of the Fifth and the Fourteenth Amendments provides the framework for testing evidence rules for constitutionality.").

169 *Id.* at 88 ("Today's constitutional doctrine holds that bad evidence may invalidate the defendant's conviction on due process grounds.").

170 *Id.* at 89 (citing Mooney v. Holohan, 294 U.S. 103, 110, 112–13 (1935)).

171 *Id.* (citing Foster v. California, 394 U.S. 441, 442–43 (1969)).

172 *Id.*

173 Medina v. California, 505 U.S. 437, 445 (1992).

what is "ranked as fundamental" is hardly scientific. Courts must "test the fundamental nature of a right within the context of that *common law* system of justice, rather than against some hypothesized 'civilized system' or some foreign system growing out of different traditions."[174] Thus, a right is fundamental if it "is necessary to an Anglo-American regime of ordered liberty."[175]

A defendant's most apparent argument would be, first, as a threshold matter, that the common law tradition has always required evidence that is explainable, bears discernible logic, and may be examined or challenged; and second, that machine learning evidence does not fit that mold because oftentimes experts cannot discern how the machine made a particular determination. Unfortunately, a defendant would have difficulty demonstrating a long-established recognition at common law that evidence must be *fully* explainable. Then, even if the Court recognized such a view, it is not clear machine learning evidence would meet that definition, as its processes, methodology, data, and assumptions can, if not fully, be *mostly* explained and understood. Machine learning does have logical, scientific, and mathematical principles; and while it does make errors, such rates of error are knowable.[176] Moreover, machine learning output would likely be introduced in the form of expert testimony,[177] meaning the defendant would have the opportunity to cross-examine an expert on the machine's capabilities and processes.

The Wisconsin Supreme Court in *Loomis* examined an issue that bears some resemblance to the above due process inquiry, though different from the issue of substantive evidence at trial. After pleading guilty to several offenses related to a drive-by shooting, Loomis appeared for his sentencing hearing,[178] and in determining his sentence, the court relied on a report generated by COMPAS, one of the most popular risk assessment tools in the United States.[179]

Loomis challenged his sentencing determination in part on due process grounds, arguing that he and the sentencing judge knew little about how the algorithm worked or the extent to which it relied on his gender in making a risk estimation.[180] In essence, he argued, it was black-box sentencing, thus

---

174 Israel, *supra* note 165, at 384 (emphasis added).

175 *Id.* (quoting Duncan v. Louisiana, 391 U.S. 145, 149 n.14 (1968)).

176 For discussions of machine learning and error rates, see *supra* Parts III and & IV.

177 *See supra* Part IV.

178 State v. Loomis, 881 N.W.2d 749, 754 (Wis. 2016).

179 Julia Angwin et al., *Machine Bias*, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing ("Northpointe's software is among the most widely used assessment tools in the country.").

180 *Loomis*, 881 N.W.2d at 765 ("Loomis asserts that because COMPAS risk scores take gender into

offending his due process right to be sentenced on the basis of accurate information.[181] The court conceded that, due to trade secret protections, neither it nor the parties understood fully how the algorithm worked or the extent to which gender was a factor.[182] Yet the court ultimately decided that because Loomis could challenge the inputs and outputs of the algorithm—that is, the data that went in and the conclusions drawn from it—he had sufficient basis to challenge the algorithm even without knowing the extent of its internal processes,[183] and thus he enjoyed due process. As for the use of gender, the court concluded that the use of gender worked to promote the accuracy of the algorithm's conclusions, which satisfied due process.[184]

Again, *Loomis* is instructive insofar as the Wisconsin Supreme Court was evaluating due process rights in the context of unknown algorithmic processes, but it was not a case of machine learning or evidence at trial. Even so, *Loomis* shows courts' reluctance to find new due process rights in black box, algorithmic evidence. When machine learning evidence is used at trial to help prove guilt beyond a reasonable doubt, courts may echo the *Loomis* decision and similarly find due process satisfied when (1) the defendant can at least challenge the data that go into the algorithm (a requirement that can be addressed with procedural rules and discovery wholly within the Court's control) and (2) the algorithm possesses some sufficient level of accuracy, which can come to light at a *Daubert* hearing on admissibility on cross-examination at trial.

---

account, a circuit court's consideration of a COMPAS risk assessment violates a defendant's due process right not to be sentenced on the basis of gender.").

181 *Id.* at 760 ("It is well-established that a defendant has a constitutionally protected due process right to be sentenced upon accurate information." (internal citations omitted)). *See also id.* ("The plurality opinion [in *Gardner*] concluded that the defendant 'was denied due process of law when the death sentence was imposed, at least in part, on the basis of information which he had no opportunity to deny or explain.'" (citing Gardner v. Florida, 430 U.S. 349, 351 (1977))).

182 *Id.* at 761 ("Northpointe, Inc., the developer of COMPAS, considers COMPAS a proprietary instrument and a trade secret. Accordingly, it does not disclose how the risk scores are determined or how the factors are weighed.").

183 *Id.* at 761–62 ("Loomis had an opportunity to challenge his risk scores by arguing that other factors or information demonstrate their inaccuracy.").

184 *Id.* at 766 ("Likewise, there is a factual basis underlying COMPAS's use of gender in calculating risk scores. Instead, it appears that any risk assessment tool which fails to differentiate between men and woman will misclassify both genders.").

B. Confrontation Under the Sixth Amendment

When the prosecution seeks to admit machine learning evidence, it is likely that the Sixth Amendment's Confrontation Clause would require an expert to testify in-person and be subject to cross examination.[185] Analogously, when other forms of machine evidence have been used in prosecution, the Supreme Court has held that, under the Sixth Amendment, the results may not be admitted without an expert subject to cross examination.

Indeed, the manner in which the Sixth Amendment requires expert witnesses to testify on drug analysis evidence may provide a framework for how machine learning experts would be required to testify in prosecutions.[186] Similar to how lab scientists are required to testify in-person and be subject to cross examination, it is likely that a machine learning expert would also have to appear in person to admit inculpatory machine learning output into evidence.

VI. THE WEIGHT OF UNEXPLAINABLE MACHINE LEARNING EVIDENCE

Parts IV and V establish that there is nothing inherently inadmissible about machine learning evidence under the Federal Rules of Evidence, the Fifth Amendment, or the Sixth Amendment. Yet, assuming the machine learning evidence is admissible, "there can be significant remaining questions about the weight and believability of the evidence."[187] Indeed, when judges rule on the admissibility of scientific evidence, they are expressly playing the role of gatekeeper and rejecting only evidence that is not the product of reliable principles and methods.[188] Rejecting scientific evidence seems to be exception rather than the norm, in keeping with the Supreme Court's observation that the Federal Rules of Evidence are construed to be liberal and permissive.[189] In fact, Professors Helland and Klick conclude "there is virtually no systematic evidence supporting the view that adoption of *Daubert*

---

185 U.S. CONST. amend. VI ("In all criminal prosecutions, the accused shall enjoy the right . . . to be confronted with the witnesses against him . . . .").

186 *See* Melendez-Diaz v. Massachusetts, 557 U.S. 305, 316–18 (2009) (reasoning that any perceived objectivity in scientific evidence does not render it immune from the Confrontation Clause); *see also* Erick J. Poorbaugh, Note, *Interfacing Your Accuser: Computerized Evidence and the Confrontation Clause Following* Melendez-Diaz, 23 REGENT U. L. REV. 213, 229 (2010) ("Although the Supreme Court in *Melendez-Diaz* stated that the 'witnesses' in that case were 'the analysts,' it did not specify which of the analysts must testify (or whether they all must testify).").

187 Imwinkelried, *supra* note 128 at 118.

188 *See* Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 597 (1993) (referring to a "gatekeeping role for the judge").

189 *See* Imwinkelried, *supra* note 128, at 118 ("[T]he Court . . . characterized the general spirit of the Federal Rules as 'liberal' and 'permissive.').

makes any difference at all" in keeping "junk science" out of the courtroom.[190] Thus, for the oft-admitted "shaky but admissible evidence," as Justice Blackmun put it in *Daubert*, "[v]igorous cross-examination, presentation of contrary evidence, and careful instruction on the burden of proof" are the principal tools for swaying the trier of fact.[191]

For three reasons, these burdens for the counsel who aims to discount the persuasiveness of the "shaky but admissible evidence" are likely greater in the criminal context than in the civil liability context that Justice Blackmun was presiding over in *Daubert*. First, as practitioners have observed, "the appellate courts appear to be more willing to second-guess trial court judgments on the admissibility of purported scientific evidence in civil cases than in criminal cases."[192] That is, courts seem to be less demanding or rigorous of scientific evidence in criminal cases, which, to the dissatisfaction of many, has often included "shaky" evidence like handwriting analysis, hair comparisons, fingerprint examinations, firearms identifications, bitemark analysis, and intoxication testing.[193] Most challenges to admissibility of these types of evidence have been unsuccessful, even while exposing "the lack of empirical support for many commonly employed forensic techniques."[194] Second, machines are often afforded a presumption of reliability that can make them unduly persuasive to a lay person.[195] And third, once a trial court rules that evidence is admissible, the appellate courts are highly deferential to that decision, and only review under an abuse of discretion standard.

Given that state of affairs, losing the admissibility battle puts considerable onus on trial counsel to persuade the trier of fact to discount the weight that the evidence should be assigned. Jurors might be cautious to assign much weight to machine learning evidence because of its peculiar property that it is often not explainable.

That is, even if one has cleared the above trade secret hurdles, probed the data, and examined the source code, often no one can explain how or why a machine learning algorithm reached a particular result, which may (or may not) significantly reduce the weight it is assigned by the trier of fact. Regardless of machine learning's potential use in litigation specifically, this issue is severe because machine learning could be useful in many areas that

190 Eric Helland & Jonathan Klick, *Does Anyone Get Stopped at the Gate? An Empirical Assessment of the* Daubert *Trilogy in the States*, 20 SUP. CT. ECON. REV. 1, 32 (2012).

191 *Daubert*, 509 U.S. at 596.

192 NAT'L RES. COUNCIL, STRENGTHENING FORENSIC SCIENCE IN THE UNITED STATES: A PATH FORWARD 11 (2009).

193 *Id.* at 94, 104, 107–08, 117–18.

194 1 COURTROOM CRIMINAL EVIDENCE § 614 (6th ed. 2016).

195 Kroll et al., *supra* note 4, at 680 ("[D]ecisions made by computers may enjoy an undeserved assumption of fairness or objectivity.").

require ex post rationale and explanation.[196] Given the present state of the technology, it is foreseeable that when machine learning begins to produce substantive evidence in litigation, an expert witness on the stand—perhaps even the individual who created the machine learning algorithm at issue—would not be explain how exactly it yielded the inculpatory results. This unexplainability may even be machine learning's most concerning feature to jurors and lead them to discount the weight it should be afforded.

A. Examples of Inexplicable Machine Learning

To illustrate how this problem manifests, consider the earlier example of a programmer who is training a machine learning algorithm to recognize her mother's face in photographs.[197] As explained, the algorithm could be identifying the mother by means that humans do, such as recognizing the collection of features in the height and width of the face, shape of the head and hair, and so on.[198] But sometimes the machine might establish correlations and rules that are not apparent at first glance or that humans would not use. For instance, if the machine has only ever learned from images in which the mother was photographed with flash on, the machine may use the brightness of the image as a basis to identify the mother, and with more weight than any attribute about her face.[199] If this were the case, when the machine later must confront an image of the mother in which she was not photographed with flash, the machine might not be able to identify her (a false negative), even though humans would not be confused by such a situation. Conversely, the machine might mistakenly identify as the mother an entirely different woman who was photographed with flash (a false

---

196 These areas include, but are not limited to, national security. *See* Cliff Kuang, *Can A.I. Be Taught to Explain Itself?*, N.Y. TIMES (Nov. 21, 2017), https://www.nytimes.com/2017/11/21/magazine/can-ai-be-taught-to-explain-itself.html (quoting a national security analyst's legal need for explainable AI decisions, "If I'm going to sign off on a decision, I need to be able to justify it.").

197 For an initial discussion of this example, see *supra* Section II.B. This particular example is hypothetical and is only offered to illustrate the general issue that machines can, and do, learn unforeseen rules. It is not from a specific study.

198 *See* Murphy, *supra* note 56 ("The software extracts information from thousands of facial data points, including nose width, mustache shape, eyebrows, corners of the mouth, hairline and even aspects of the face we don't have words for. It then turns the faces into numbers.").

199 Machine learning researchers are well aware of this issue and have confronted it in a variety of contexts:

> Tomaso Poggio, the director of M.I.T.'s Center for Brains, Minds and Machines, offered a classic parable used to illustrate this disconnect. The Army trained a program to differentiate American tanks from Russian tanks with 100 percent accuracy. Only later did analysts realize that the American tanks had been photographed on a sunny day and the Russian tanks had been photographed on a cloudy day. The computer had learned to detect brightness.

Id.

positive). In such a scenario, the machine learned a correlation that was undoubtedly accurate within the universe of data it was initially shown, but not one that would be reliable for all varying situations. This is a common problem with the rules that machines learn.[200]

Whether the machine is deducing obvious rules (like facial attributes) or non-obvious and potentially unreliable rules (like brightness in an image) is impossible to predict ex ante and discovering what rules the machine has deduced sometimes requires considerable extra research for the programmer. Indeed, what rules and correlations the machine deduces may forever remain a mystery.

This principle was at work in a recent Stanford University study that aimed to build a machine learning algorithm that could analyze a person's face and determine that person's sexual orientation.[201] Specifically, researchers compiled images of 75,000 users' faces from various dating sites and used the profiles' self-reported gay or straight identification to train the algorithm.[202] From this pool of data, the algorithm focused on 35,000 images of 15,000 users to learn a set of correlations between the content of the images and the labels "gay" and "straight." Later, in a test set of different images that the machine had never seen before, the algorithm would make its best guess.[203] The program was remarkably accurate at determining straight versus gay men, at eighty-one percent accuracy, and slightly less accurate at sorting gay versus straight women, at seventy-one percent.[204] Meanwhile, the machine was far more accurate than humans, who only correctly determined male sexual orientation sixty-one percent of the time and that of women fifty-four percent of the time.[205]

---

200 *See* Hubert L. Dreyfus & Stuart E. Dreyfus, *What Artificial Experts Can and Cannot Do*, 6 AI & SOC'Y 18, 21 (1992) (further explaining the same American-Russian tank example). The issue can even lead to needless deaths in emergency situations:

> [One algorithm created to better triage emergency room patients] seemed to show that asthmatics with pneumonia fared better than the typical patient. This correlation was real, but the data masked its true cause. Asthmatic patients who contract pneumonia are immediately flagged [by doctors] as dangerous cases; if they tended to fare better, it was because they got the best care the hospital could offer. A dumb algorithm, looking at this data, would have simply assumed asthma meant a patient was likely to get better—and thus concluded that they were in less need of urgent care. . . . The story of asthmatics with pneumonia eventually became a legendary allegory in the machine-learning community.

Kuang, *supra* note 196.

201 *See* Murphy, *supra* note 56.

202 Yilun Wang & Michael Kosinski, *Deep Neural Networks Are More Accurate than Humans at Detecting Sexual Orientation from Facial Images*, 114 J. PERSONALITY & SOC. PSYCHOL. 246, 248 (2018).

203 *Id.* at 249.

204 *Id.* at 250.

205 *Id.* at 253.

As to *how* the algorithm was making its relatively accurate determinations, the researchers could only speculate. One of their hypotheses was that the levels of different hormones in gay versus straight users (the prenatal hormone theory of sexual orientation[206]) might have manifested some minute differences in their respective facial structures, differences unseen by the human eye but detectable by the algorithm.[207] One researcher explained, "[h]umans might have trouble detecting these tiny footprints that border on the infinitesimal" but "[c]omputers can do that very easily."[208] With respect to men, presence of facial hair or baseball caps also probably played a role for some images.[209] In other similar tests, researchers found that the images of gay men usually were better quality and had better lighting, and the algorithm may have used that as the basis to conclude the sexual orientation.[210] Yet, the author of the Stanford study concedes that he could not say with certainty how the algorithm made its determinations.[211]

B. Why Machine Learning Is Unexplainable

Machine learning is often unexplainable because of the sheer number of data points involved and "avalanche of statistical probability" involved.[212] Many techniques are at play or constantly being developed, and choosing among them can be the whim or preference of the programmer. "The sheer proliferation of different techniques, none of them obviously better than the others, can leave researchers flummoxed over which one to choose. Many of the most powerful are bafflingly opaque; others evade understanding because they involve an avalanche of statistical probability."[213]

Responding to that deficiency is an entirely new subfield of machine learning research, dubbed "xAI," for "explainable AI."[214] The Defense

---

206 The prenatal hormone theory is commonly circulated but controversial in both the scientific and LGBT advocacy communities. *See* Louis Hoffman & Justin Lincoln, *Science, Interpretation, and Identity in the Sexual Orientation Debate: What Does Finger Length Have to Do With Understanding a Person?*, 56 PSYCCRITIQUES, Apr. 13, 2011, http://psqtest.typepad.com/blogPostPDFs/201103880_psq_56-15_scienceInterpretationAndIdentifyInTheSexualOrientationDebate.pdf (reviewing SIMON LEVAY, GAY, STRAIGHT, AND THE REASON WHY: THE SCIENCE OF SEXUAL ORIENTATION (2011) and commenting on the controversy surrounding the introduction of science into politically charged areas such as sexual identity).

207 Wang & Kosinski, *supra* note 202, at 246.

208 Kuang, *supra* note 196.

209 *See* Murphy, *supra* note 56 (showing that the algorithm looked at factors like grooming habits).

210 *Id.*

211 Kuang, *supra* note 196.

212 *Id*.

213 *Id.*

214 *See, e.g.*, Mark G. Core et. al., *Building Explainable Artificial Intelligence Systems*, 21 PROC. NAT'L CONF. ARTIFICIAL INTELLIGENCE 1766, 1766 (2006) (setting forth a modular and generic architecture for

Department's Defense Advanced Research Projects Agency ("DARPA") is currently conducting research into how AI technologies can explain their decision-making processes, though the field is still in its infancy.[215] Thus, for the foreseeable future, any machine learning output that is admitted into evidence bears a substantial likelihood that it will be unexplainable.

C. Comparing Machine Learning to Current Unexplainable Evidence

Few analogs exist to this problem in other forms of evidence used at trial. When scholars write about "black boxes" and evidence, they typically mean to highlight the fact that lay jurors do not fully understand how the device works—the implicit assumption is that experts do. But as machine learning exists now, that assumption is faulty, since experts often cannot fully account for machine learning determinations, in spite of the machine's demonstrable accuracy for certain tasks.

Professor Rich analogizes algorithms to another black box with which courts are very familiar in the Fourth Amendment sphere: drug dogs. With drug sniffing dogs just as in algorithms, "we know the inputs, and we receive the outputs, but we cannot fully understand how the internal mechanism works."[216] Professor Rich argues that treating algorithms in the way that we do dogs will allow "courts and police to ignore what they are ill-equipped to evaluate," namely, how an algorithm works, and focus only on the accuracy of its outcomes.[217] Professor Rich's analogy is limited, however, to how algorithms may be used to develop reasonable suspicion or probable cause, and does not speak to admissibility or weight.

A closer analogy might be "super recognizers,"[218] or humans with the uncanny ability to recognize even the most blurry or corrupted images of a face to aid in investigations.[219] The super recognizer essentially exploits two

---

explaining the behavior of simulated entities); David Gunning, *Explainable Artificial Intelligence (XAI)*, DARPA/I20, https://www.darpa.mil/program/explainable-artificial-intelligence (last visited Jan. 5, 2018) (arguing that explainable AI is necessary if users are to understand, trust, and effectively manage new AI).

215 Gunning, *supra* note 214.

216 Rich, *supra* note 11, at 912.

217 *Id.* at 919.

218 *See* Richard Russell et al., *Super-recognizers: People with Extraordinary Face Recognition Ability*, 16 PSYCHONOMIC BULL. & REV. 252, 252 (2009) (defining the term super-recognizers).

219 Anna K. Bobak & Sarah Bate, *Superior Face Recognition: A Very Special Super Power*, SCI. AM.. (Feb. 2, 2016), https://www.scientificamerican.com/article/superior-face-recognition-a-very-special-super-power/ ("London's Metropolitan Police have created a super-recogniser unit that is used to spot criminals in a crowd or within CCTV footage. . . . It's easy to spot other potential roles for super-recognisers—issues of national security are currently paramount, and they may spot wanted or missing people more readily than typical officers.").

capabilities in tandem. First is the ability to remember a face after de minimis exposure, and second is the ability to recognize it in subsequent situations. How exactly a super recognizer memorizes the face with such precision is a mystery. One hypothesis was that super recognizers holistically evaluate a face differently than the average layperson, though that possibility was debunked in subsequent research.[220] Another hypothesis was that super recognizers spend longer than average looking at the eyes,[221] though subsequent research showed they actually spend more time focusing on the nose and center of the face.[222] In any case, the super recognizers cannot themselves articulate how their ability works, nor can other researchers provide a comprehensive explanation. The super recognizers are a human black box.

If a super recognizer were to testify at trial on the basis of her abilities and give inculpatory evidence, it would perhaps give some kind of analog to how a testifying expert cannot account for how a machine learning algorithm made the same kind of facial recognition determination. At present, however, there is no documented use of a super recognizer testifying at trial, and thus no instance of how a judge rationalized the admissibility of one's testimony, to say nothing of how jurors assigned it weight and credibility.[223]

D. Jurors' Trust in Unexplainable Machine Learning Evidence

It is an entirely open question the extent to which, in open court, jurors would trust the validity of unexplainable machine learning evidence. Indeed, this question is ripe for empirical research by psychologists and legal scholars of scientific evidence.

Developers understand that the extent to which a person trusts a machine in everyday life is highly variable and context-dependent. Outside the courtroom, an individual's trust in a machine ranges from none or little (for a variety of reasons, one of which is often *because it is a machine*[224]), to passive

---

220 *Id*. ("[O]nly modest links have been reported between face recognition ability and holistic processing skills, suggesting other factors may be at play.").

221 *Id.* ("Eye-tracking technology has frequently been used by psychologists to identify the regions of the face that are particularly informative in face recognition. Typical people tend to focus on the eyes, suggesting they carry important information about facial identity.").

222 *Id.* ("[S]uper-recognisers spent more time viewing the nose. These findings challenge existing conclusions, suggesting that it is the centre[sic] of the face, rather than the eye region, that is optimal for facial identity recognition.").

223 Gary Edmond & Natalie Wortley, *Interpreting Image Evidence: Facial Mapping, Police Familiars and Super-Recognisers in England and Australia*, 3 J. INT'L & COMP. L. 473, 492 (2016) ("So far there are no reported cases involving police super-recognisers as witnesses.").

224 Berkeley J. Dietvorst et al, *Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err*, 144 J. EXPERIMENTAL PSYCHOL. 1, 10 (2014) ("The results of five studies show that seeing

trust in machines without so much as a second thought[225] (when the machine is functioning, that is[226]), and even up to great affirmative trust in a machine *because it is a machine*.[227] Researchers find trust in machines to be highly variable and influenced by different factors like belief about the functionality of the technology, belief that the technology is helpful, and belief that the technology is reliable.[228] Often, trust is dependent on mere different presentations of a technology—such as the inclusion of anthropomorphic qualities, for example[229]—and not by the capabilities of the machine itself.

To be sure, trust itself means different things in different contexts, and the meanings may not necessarily translate from one situation to another. For example, consider voir dire, where a prosecutor hoping to use machine learning output at trial may question potential jurors about their trust in technology in specific situations. It would be risky for the prosecution to assume, say, that a juror's self-reported "trust in technology" would necessarily indicate trust in black-box machine output to prove a defendant's guilt beyond a reasonable doubt. Inside the courtroom, how jurors will respond to machine learning output is very difficult to predict. As Professor Roth summarizes, "juries might irrationally defer to the apparent objectivity

---

algorithms err makes people less confident in them and less likely to choose them over an inferior human forecaster.").

225 *See* Paul Robinette et al., *Overtrust of Robots in Emergency Evacuation Scenarios*, 11 ACM/IEEE INT'L CONF. HUMAN ROBOT INTERACTION 101, 104 (2016) (finding that in a simulated emergency situation 100% of human test subjects followed a guide robot to safety through the hallways of a building even when the robot led in directions opposite posted exit signs); *see also id.* ("Eighty-five percent of participants indicated that they would follow the robot in a future emergency.").

226 *See* Davide Salanitri et al., *Relationship Between Trust and Usability in Virtual Environments: An Ongoing Study*, *in* HUMAN COMPUTER INTERACTION: DESIGN AND EVALUATION 49, 50 (Masaaki Kurosu ed. 2015) ("A low level of usability could compromise the users' interaction with a product, thus affecting the individual's trust in the technology . . . .").

227 *See* Jennifer Saranow, *Steered Wrong: Drivers Trust GPS Even to a* Fault, WALL STREET J. (Mar. 18, 2008, 11:59 PM), https://www.wsj.com/articles/SB120578983252543135 ("If your GPS device told you to drive off a clib would you do it? Norman Sussman nearly did."); *see also* The Office US, *Michael Drives into a Lake—The Office US*, YOUTUBE (Aug. 8, 2017) https://www.youtube.com/watch?v=DOW_kPzY_JY (satirizing the enormous trust that drivers have in turn-by-turn GPS navigation by depicting a driver intentionally driving into a lake because "maybe it's a shortcut" and "the machine knows").

228 Salanitri, *supra* note 226, at 50 ("[T]rust in a technology reflects at least three main beliefs about the attributes of a technology: (i) belief about the functionality of the product, which refers to the capability of a technology to perform specific tasks; (ii) belief that the technology is helpful . . . (iii) Belief that the technology is reliable, hence, the perception that a technology works properly.").

229 For instance, even though the two technologies provide the same service, first-time passengers report more trust in Tesla's self-driving technology than Google's, owing in part Tesla's incorporation of human-like qualities a voice and a name. Google's, meanwhile, has no anthropomorphic qualities. *See* Walter Frick, *Tesla, Autopilot, and the Challenge of Trusting Machines*, HARV. BUS. REV. (July 11, 2016), https://hbr.org/2016/07/tesla-autopilot-and-the-challenge-of-trusting-machines.

of machines, or reject machine sources because of an irrational mistrust of machines' apparent complexities, even when the sources are highly credible."[230]

Additionally, inextricably linked to the credibility of the machine is the credibility the jurors extend to the testifying expert him- or herself. That human credibility would likely affect credibility that jurors would extend to the underlying machine, especially as the scientific evidence at issue is particularly complex for laypeople.[231] In that case, the prosecution or defense would surely already be familiar with the usual tactics to use to attack the expert's credibility. These tactics include choosing and preparing an expert who (1) appears to lack bias, (2) bears impressive credentials, (3) displays "a pleasant personality," (4) can present "a clear, objective, focused, not overly long presentation that utilizes diagrams and models," (5) uses lay terms, (6) demonstrates knowledge in the area of expertise, (7) gives testimony that is "complete, consistent, and not too complex," and (8) shows familiarity with the instant case.[232] As for attacking credibility, studies seem to show the classic methods are effective. Jurors report less credibility for experts that seem like "hired guns," as in experts that are highly paid and bear sterling credentials.[233] Jurors also distrust witnesses that offer inconsistent statements between depositions and trial testimony.[234]

---

230 Roth, *Testimony*, *supra* note 17, at 2038 (2017); *see also* Hon. Donald E. Shelton et. al., *An Indirect-Effects Model of Mediated Adjudication: The* CSI *Myth, the Tech Effect, and Metropolitan Jurors' Expectations for Scientific Evidence*, 12 VAND. J. ENT. & TECH. L. 1, 8 (2009) ("Data in the Washtenaw County and Wayne County studies have demonstrated high expectations and demands for scientific evidence among jurors. Other scholars and researchers have found similarly high expectations and regard for scientific evidence by jurors.").

231 Researchers typically reason that jurors evaluate complicated expert testimony either by (1) evaluating the logic of the testimony itself and trying to understand the underlying science (i.e., "central processing"), or when that is difficult, (2) reverting to shortcuts, heuristics, or other means of evaluating the testimony (i.e., "peripheral processing"), which includes things like the credentials of the expert, how much the expert has been paid, or even the expert's manner of speech or appearance. *See* Sanja Kutnjak Ivkovic & Valerie P. Hans, *Jurors' Evaluations of Expert Testimony: Judging the Messenger and the Message*, 28 LAW & SOC. INQUIRY 441, 448 (2003) ("Cooper and Neuhaus concluded that jurors shifted from central to peripheral processing under cognitively challenging conditions.").

232 *Id.* at 458.

233 *Id.* at 448 ("[M]ock jurors who heard testimony of a highly paid expert with high credentials—potentially fitting the profile of a hired gun—rated the expert as less likable, less believable, less trustworthy, less honest, and more annoying . . . .").

234 *Id.* at 473 ("This shows that the common litigator tactic of pointing to differences between deposition and trial testimony can be effective in decreasing credibility.").

CONCLUSION

Machine learning is already in our email clients, our web applications, our law firms, and our government's regulatory agencies. It will soon arrive in our courtrooms, too. When it does, it will only be the latest in a long line of machine evidence that is admitted in spite of the risk of error it carries. While machine learning poses some risks under Federal Rule of Evidence 702—namely that its data must be appropriately compiled and relevant to the given defendant—nothing in the Federal Rules of Evidence inherently bars machine learning output as a form of evidence. For its part, the Constitution and relevant precedent permit machine learning evidence even in spite of its unexplainability, even if the Sixth Amendment merely requires that the evidence be introduced with expert testimony.