PEPPERDINE UNIVERSITY

Pepperdine Law Review

Volume 51 | Issue 4 Article 1

4-15-2024

Machines Like Me: A Proposal on the Admissibility of Artificially Intelligent Expert Testimony

Andrew W. Jurs

Scott DeVito

Machines Like Me:

A Proposal on the Admissibility of Artificially Intelligent Expert Testimony

Andrew W. Jurs* & Scott DeVito**

ABSTRACT

With the rapidly expanding sophistication of artificial intelligence systems, their reliability, and cost-effectiveness for solving problems, the current trend of admitting testimony based on artificially intelligent (AI) systems is only likely to grow. In that context, it is imperative for us to ask what rules of evidence judges today should use relating to such evidence. To answer that question, we provide an in-depth review of expert systems, machine learning systems, and neural networks. Based on that analysis, we contend that evidence from only certain types of AI systems meet the requirements for admissibility, while other systems do not. The break in admissible/inadmissible AI evidence is a function of the opaqueness of the underlying computational methodology of the AI system and the court's ability to assess that methodology. The admission of AI evidence also requires us to navigate pitfalls including the difficulty of explaining AI systems' methodology and issues as to the right to confront witnesses. Based on our analysis, we offer several policy proposals that would address weaknesses or lack of clarity in the current system. First, in light of the long-standing concern that jurors would allow expertise to overcome their own assessment of the evidence and blindly agree with the "infallible" result of advanced-computing AI, we propose that jury

* Richard M. and Anita Calkins Distinguished Professor of Law, Drake University Law School. Thanks to Katie, Clara, and Milo.

** Associate Professor of Law, Jacksonville University College of Law; Ph.D. University of Rochester, J.D. University of Connecticut. The author thanks the College of Law for its support.

instruction commissions, judicial panels, circuits, or other parties who draft instructions consider adopting a cautionary instruction for AI-based evidence. Such an instruction should remind jurors that the AI-based evidence is solely one part of the analysis, the opinions so generated are only as good as the underlying analytical methodology, and ultimately, the decision to accept or reject the evidence, in whole or in part, should remain with the jury alone. Second, as we have concluded that the admission of AI-based evidence depends largely on the computational methodology underlying the analysis, we propose for AI evidence to be admissible, the underlying methodology must be transparent because the judicial assessment of AI technology relies on the ability to understand how it functions.

TABLE OF CONTENTS

I. PITFALLS AND PROPOSALS FOR ARTIFICIALLY INTELLIGENT EXPERT EVIDENCE

Artificial intelligence (AI) has been a near-daily obsession in the media recently, as society begins to grapple with the combined power of today's processors and the explosion of data available in a variety of fields.[1] AI has been hailed as a tool with nearly limitless potential for societal benefits, but there is also the potential for AI to lead to significant negative consequences.[2] In the legal field, AI has been used in different ways, from e-discovery[3] to assessment of flight risk for bail,[4] but the usage as evidence is only beginning

1. *See* David Ingram, *Trying to Make Sense of Artificial Intelligence? Here's Your Guide*, NBC NEWS (May 16, 2023, 6:09AM), https://www.nbcnews.com/tech/innovation/ai-explain-openai-chatgpt-how-to-rcna77889. A Lexis search for *New York Times* headlines in the past thirty days reveals fifty-five articles with the term "AI" in the headline, plus additional articles with "algorithm" and "artificial intelligence," including articles on regulation of artificial intelligence, AI's ability to evaluate brain activity, and its potential dangers. *See, e.g.*, Lina M. Kahn, *The U.S. Needs to Regulate A.I.*, N.Y. TIMES, May 6, 2023, at A21; Oliver Whang, *A.I. to Read Your Mind Is Up Next*, N.Y. TIMES, May 2, 2023, at B1; Cade Metz, *If Some Dangers Posed by A.I. Are Already Here, Then What Lies Ahead?*, N.Y. TIMES, May 8, 2023, at B5. A search with the *Washington Post* revealed a related level of interest and similar topics. *See, e.g.*, Yan Wu & Sergio Peçanha, *Type in Your Job to See How Much AI Will Affect It,* WASH. POST (May 9, 2023, 9:17 AM), https://www.washingtonpost.com/opinions/interactive/2023/ai-artificial-intelligence-jobs-impact-research/; Danielle Allen, *The Next Level of AI Is Approaching. Our Democracy Isn't Ready.*, WASH. POST (Apr. 26, 2023, 6:30 AM), https://www.washingtonpost.com/opinions/2023/04/26/artificial-intelligence-democracy-danielle-allen/.

2. *Compare* NAT'L SCI. & TECH. COUNCIL COMM. ON TECH., EXEC. OFFICE OF THE PRESIDENT, PREPARING FOR THE FUTURE OF ARTIFICIAL INTELLIGENCE 1, 5 (2016) ("One area of great optimism about AI and machine learning is their potential to improve people's lives by helping to solve some of the world's greatest challenges and inefficiencies. . . . Artificial Intelligence (AI) has the potential to help address some of the biggest challenges that society faces.") *with* Metz, *supra* note 1, at B5 (cataloging some risks of AI deployment, from job loss to loss of societal control, and mentioning a letter asking for a moratorium on AI development was written by AI experts concerned that AI could cause harm to society).

3. *See* Catrina Denvir et al., *The Devil in the Detail: Mitigating the Constitutional & Rule of Law Risks with the Use of Artificial Intelligence in the Legal Domain*, 47 FLA. ST. U. L. REV. 29, 61–62 (2019) (discussing the use of AI for e-discovery purposes, including different methodologies of analysis and commercial software available in the field); *see also* Tammy Pettinato Oltz, *Educating Robot-Proof Attorneys*, 97 N.D. L. REV. 185, 199–200 (2022) (discussing how AI has made the discovery process faster and cheaper).

4. *See* Ngozi Okidegbe, *Discredited Data*, 107 CORNELL L. REV. 2007, 2027–32 (2022) (reviewing the use and construction of pretrial release algorithms); *see also* Arthur Rizer & Caleb Watney, *Artificial Intelligence Can Make Our Jail System More Efficient, Equitable, and Just*, 23 TEX. REV. L. & POL. 181, 191–94 (2018) (discussing the history and current applications of risk assessment in the pretrial and jail systems).

and has only recently been the focus of court and commentator analysis.[5]

Consider for a moment the use of forensic evidence in the criminal justice system starting in the 1930s.[6] Initially, these developments were solely seen as a positive good, a net increase in reliability and a more "scientific" method of finding guilt.[7] For decades, judicial assessments of these methods reflected this utopian vision, rarely invoking skepticism in admitting a wide range of forensic evidence.[8] Yet recent developments, specifically the amazing specificity of nuclear DNA, have shown that in many fields, the lenient admissibility approach was a grave error.[9] Too many convictions have been found to have rested on a shaky if not fundamentally flawed foundation,[10] and a wide variety of evidentiary techniques once thought to be nearly infallible (e.g., microscopic hair comparison, bullet-lead evidence, footwear analysis, and

5. *See, e.g.*, JAMES E. BAKER ET AL., AN INTRODUCTION TO ARTIFICIAL INTELLIGENCE FOR FEDERAL JUDGES 24 (2023), https://permanent.fdlp.gov/gpo195237/An_Introduction_to_Artificial_Intelligence_for_Federal_Judges.pdf (explaining that judges will need to scrutinize the reliability of evidence generated by AI); Paul W. Grimm et al., *Artificial Intelligence as Evidence*, 19 NW. J. TECH. & INTELL. PROP. 9, 48 (2021) (discussing how the validity and reliability of an AI system must be tested to determine if the algorithm is trustworthy); Andrea Roth, *Machine Testimony*, 126 YALE L.J. 1972, 2021–22 (2017) (stating how the use of AI in litigation raises reliability issues not addressed by current law on evidence).

6. JOSEPH PETERSON ET AL., NAT'L INST. OF JUSTICE, THE ROLE AND IMPACT OF FORENSIC EVIDENCE IN THE CRIMINAL JUSTICE PROCESS 14 (2010), https://www.ojp.gov/pdffiles1/nij/grants/231977.pdf (discussing the emergence of forensic evidence in the 1930s).

7. *See* Rachel E. Barkow, *Prosecutorial Administration: Prosecutor Bias and the Department of Justice*, 99 VA. L. REV. 271, 293 (2013) (explaining the creation of the FBI Crime Lab in 1932 to develop forensic science as "a key part of the agency's mission"); John F. Fox, Jr., *The Birth of the FBI's Technical Laboratory—1924 to 1935*, FBI, https://www.fbi.gov/history/history-publications-reports/the-birth-of-the-fbis-technical-laboratory1924-to-1935 (last visited Feb. 2, 2024) (stating "the application of science to criminal investigations was becoming a Bureau priority," and led to the creation of the FBI crime lab in 1932). This is not to say that forensic disciplines did not exist prior to the 1930s. *See, e.g.*, Jennifer L. Mnookin, *Fingerprint Evidence in an Age of DNA Profiling*, 67 BROOK. L. REV. 13, 17–30 (2001) (reviewing the history and admissibility of fingerprint evidence in the early twentieth century).

8. *See* DAVID FAIGMAN ET AL., MODERN SCIENTIFIC EVIDENCE: THE LAW & SCIENCE OF EXPERT TESTIMONY § 29:2, Westlaw (database updated Dec. 2023) (calling courts' approach to prosecutorial forensic evidence "casual acceptance," and stating it led to a "century of easygoing admission").

9. *See* COMM. ON IDENTIFYING THE NEEDS OF THE FORENSIC SCI. CMTY. ET AL., NAT'L RESEARCH COUNCIL, STRENGTHENING FORENSIC SCIENCE IN THE UNITED STATES: A PATH FORWARD 4 (2009) [hereinafter NRC REPORT] (finding that DNA analysis has shown that faulty forensics led to "wrongful convictions of innocent people").

10. *Id.* at 7 ("With the exception of nuclear DNA analysis, however, no forensic method has been rigorously shown to have the capacity to consistently, and with a high degree of certainty, demonstrate a connection between evidence and a specific individual or source.").

bitemarks) have either been debunked or shown to have limits inconsistent with their long-standing usage and admissibility.[11] These decisions can have catastrophic consequences both in individual cases and to the criminal justice system as a whole.[12]

AI is currently in the "irrational optimism" stage, and utopian visions of computer-aided improvements to criminal justice offer a tempting view of a future unencumbered by the messy complications of forensics.[13] As this Article argues below, succumbing to the visions of the prophets of unbounded optimism would be a grave error.

Instead, this Article asserts that AI has the potential to offer some benefits as an adjunct to the decision-making process but must be constrained within limits. This Article examines the issue of using AI as an expert, considers the use, misuse, and limits of AI technology as admissible evidence in courtrooms, and offers a prescription for courts to use when facing admissibility challenges in litigation today and in the near future. These limits are based in both statutory or rule-based authority, such as reliability limits under Federal Rule of Evidence 702, and in constitutional protections necessary for the legitimacy of the system to continue, such as the right of confrontation or equal protection limits.

At a fundamental level, the admissibility of evidence from an AI expert, under our current rules of evidence, is dependent on the type of AI system at play.[14] In essence, the judge must determine how the AI functions in order to assess whether its output is admissible.[15] AI has, in recent years, become a

11. *See* PRESIDENT'S COUNCIL OF ADVISORS ON SCI. & TECH., EXEC. OFFICE OF THE PRESIDENT, FORENSIC SCIENCE IN CRIMINAL COURTS: ENSURING SCIENTIFIC VALIDITY OF FEATURE-COMPARISON METHODS 3, 13 (2016) (explaining several fields that have been shown to be flawed, including microscopic hair comparison, bullet-lead evidence, footwear analysis, and bitemarks, before a comprehensive review of each field).

12. *See, e.g.*, David Grann, *Trial by Fire*, NEW YORKER (Aug. 31, 2009), https://www.newyorker.com/magazine/2009/09/07/trial-by-fire (telling a horrifying example of forensic evidence error in a single case). Regarding the compounding number of errors within the system, the Innocence Project alone has been responsible for hundreds of exonerations. *Explore the Numbers: Innocence Projet's Impact*, INNOCENCE PROJECT, https://innocenceproject.org/exonerations-data/ (last visited Mar. 7, 2024).

13. *See supra* note 2 and accompanying text (regarding the irrational optimism of AI development).

14. *See* Grimm et al., *supra* note 5, at 9, 97 (discussing how the types and uses of AI can affect the validity and reliability of AI evidence).

15. *See id.* at 105 (noting that judges will be "ill equipped" to address evidentiary issues unless they have "at least a rudimentary understanding of what AI is, how it operates, scientific and statistical

"catch-all" for a variety of computational tools and smart technology, but to understand what is and is not permissible as evidence, one must examine the program design and method in depth.[16] Using the criteria of transparency and explainability, this Article proposes that only sufficiently transparent and explainable systems (open-box AI systems) should be permissible as evidence in courtrooms today.[17] For if closed-box technology can be admitted, we truly will have jumped into the "Trial by Machine" era, where society has ceded control to larger forces in a way warned about by writers like George Orwell and Aldous Huxley.[18]

To assess these issues, Part II will first examine the current and likely near-term use of AI technologies in law enforcement or forensics more generally and identify those most likely to raise admissibility questions in court. We will then, in Part III, assess the different computation methods that underlie all these AI methods, describing categories or groupings based on levels of opacity or openness. Once established, we will use these categories in Part IV to create a road map for judicial consideration of expert substitution AI in litigation, suggesting the limits to admissibility in both civil and criminal cases. Finally, in Part V, we will offer thoughts of issues to consider as AI develops further, to preserve the legitimacy of the criminal justice system.

AI is nothing more than a tool, and as such, can be used both for good and for evil.[19] While we acknowledge the seductive power of a utopian vision of a world without messiness due to AI usage, we believe that the proper time to consider limiting use of the tool is before implementation lest we end up decades later regretting its widespread use and limiting it like with forensic fields. AI as expert evidence must be limited in court. In this Article, we provide a road map for judges on when and why to admit such evidence in current and future litigation. These commonsense measures based on current doctrine

evaluation, and the issues that need to be addressed in order to make decisions about its validity and reliability").

16. *See* discussion *infra* Part III (regarding the design and different computational methods of AI technology).

17. *See* discussion *infra* Part IV (explaining how current doctrine leads to this conclusion).

18. *See* Andrea Roth, *Trial by Machine*, 104 GEO. L.J. 1245, 1253–69 (2016); ALDOUS LEONARD HUXLEY, BRAVE NEW WORLD (1932). Andrea Roth, in a comprehensive article about the history, perils, and response to the use of technology in the judicial system, catalogued many of the ways machines have and can be used as forms of proof. Roth, *supra*.

19. Bernard Marr, *World Leaders Weigh Tech's Use of 'Good or Evil' at AI Summit*, FORBES (Nov. 3, 2023, 6:19AM), https://www.forbes.com/sites/bernardmarr/2023/11/03/world-leaders-weigh-techs-use-of-good-or-evil-at-ai-summit/?sh=63c13ed63d15 (discussing world leaders' thoughts on advantages and disadvantages of the proliferation of AI technology).

provide a safeguard under the current state of the law to balance rights and fairness as broader implications and modifications to the system will be considered. In so doing, we can preserve the legitimacy of the decision-making process against temptations for overuse and thereby preserve the legitimacy of the justice system overall during this time of transition.

II. ARTIFICIAL INTELLIGENCE—CURRENT FORENSIC USAGE AND PREVALENCE

Machine-based and machine-assisted forensic evidence has long-standing use in the legal system, from devices to measure alcohol intoxication to modern DNA evidence.[20] Artificial intelligence, on the other hand, refers to systems that engage in self-learning and creative assessment and are a relatively new phenomenon in forensics.[21] Our analysis will focus mainly on AI use for forensic evidence, which has special risks and considerations independent of those related to the wider consideration of all machine-based testimony.[22] Two fundamental questions to address at the outset are how prevalent the usage of AI is in forensics and how forensic usage for lead generation or for evidentiary use raises different considerations under current doctrine.

Before describing these categories of forensic AI though, we must acknowledge that the type, extent, and prevalence of AI forensics in law enforcement is largely opaque. The large data sets and AI systems used in implementing varying technologies are not publicly available for review, meaning that both media reporting on these issues and judicial assessment of new technologies are years behind their implementation in the field.[23] For that reason, and others, regulatory reaction to the development and

20. Roth, *supra* note 18, at 1247.

21. *What Is Artificial Intelligence (AI)?*, IBM, https://www.ibm.com/topics/artificial- intelligence (last visited Feb. 2, 2024). Stanford Professor John McCarthy first used the term "artificial intelligence" and has since defined it as "the science and engineering of making intelligent machines." *Id.* However, the exact definition of AI remains elusive. Ryan Calo, *Artificial Intelligence: A Primer and Roadmap*, 51 U.C. DAVIS L. REV. 399, 404 (2017). For example, AI textbook authors Stuart Russell and Peter Norvig have eight different definitions in four categories: thinking or acting humanly or thinking or acting rationally. STUART J. RUSSELL & PETER NORVIG, ARTIFICIAL INTELLIGENCE: A MODERN APPROACH 2 (3d ed. 2010).

22. *See* Roth, *supra* note 5, at 1972 (giving a detailed examination of the broader implications of machine testimony).

23. C. Ross Brown, *Artificial Intelligence in Law Enforcement*, LAW ENFORCEMENT BULL. (Sept. 7, 2022), https://leb.fbi.gov/articles/featured-articles/artificial-intelligence-in-law-enforcement.

implementation is also lagging.[24] Yet current media reports and judicial opinions can provide enough context to estimate both current and future forensic usage patterns.[25]

A. Use of Artificial Intelligence by Law Enforcement—Non-Evidence Use and Limits

Certain AI methods are used by law enforcement, not for specific admissible evidence for use in court, but instead to metaphorically cast a wide net of assessment in a particular area, field, or location. A review of several technologies illustrates current uses and potential future uses.

One common AI usage is to assist in predictive policing.[26] Law enforcement has analyzed crime patterns to manage resource prioritization for many years, but AI can assist the human or standard computer analysis to generate more detailed models.[27] These models can then be used for resource allocation to react to past crimes and deter future crimes.[28] They can also predict who will be at risk from these crimes,[29] not only from prior patterns but also by associations and behavior of individuals.[30]

Artificial intelligence can be used to assist or even by itself monitor audio systems for sounds, identifying gunshots and, potentially, the number and type of weapons being used.[31] In fact, the authors of a recent study predict that AI

24. Joseph Boyle, *As AI Rises, Lawmakers Try to Catch Up*, TECHXPLORE (Dec. 7, 2022), https://techxplore.com/news/2022-12-ai-lawmakers.html.

25. Alexander Eser, *Essential AI in Law Enforcement Statistics in 2024*, ZIPDO, https://zipdo.co/statistics/ai-in-law-enforcement/ (Aug. 9, 2023) (providing AI usage pattern statistics).

26. *Surveillance and Predictive Policing Through AI*, DELOITTE, https://www.deloitte.com/global/en/Industries/government-public/perspectives/urban-future-with-a-purpose/surveillance-and-predictive-policing-through-ai.html (last visited Feb. 2, 2024) (discussing predictive policing across different regions).

27. *See* JAMES REDDEN ET AL., CRIMINAL JUSTICE TESTING & EVALUATION CONSORTIUM, ARTIFICIAL INTELLIGENCE APPLICATIONS IN LAW ENFORCEMENT (2020), https://cjtec.org/files/5f5f94aa4c69b.

28. *See, e.g.*, Ethan Baron, *Predictive Policing Using AI Tested by Bay Area Cops*, GOV'T TECH. (Mar. 11, 2019), https://www.govtech.com/public-safety/predictive-policing-using-ai-tested-by-bay-area-cops.html.

29. Chritopher Rigano, *Using Artificial Intelligence to Address Criminal Justice Needs*, 280 NAT'L INST. JUST. J. 37, 43 (2018).

30. *Id.* at 44.

31. *See* Alex Morehead et al., *Low Cost Gunshot Detection Using Deep Learning on the Raspberry*

could use microphone arrays to triangulate the location of the shots, assisting law enforcement in the deployment of officers in response.[32]

Police can use AI to monitor social media, whether to look for potential threats or to create profiles of potential suspects from social media.[33] Threat assessment uses AI to monitor public postings in social media for signs of overt threats, allowing officers to react to, investigate, or mitigate the potential threat.[34] Social media monitoring can also evaluate the entirety of a person's social media usage for clues about their overall social profile, assisting law enforcement in lead generation.[35]

Video and image analysis is an AI technology that can be used for large-scale surveillance to assess patterns of conduct, identify characteristics of individuals, and generate leads to pursue by other means.[36] To the extent it has the potential for use in individual cases, the analysis would fall squarely under the forensic evidentiary rules considered below.[37] Large-scale deployment of

Pi, 2019 IEEE INT'L CONF. ON BIG DATA, Dec. 2019, at 3038, https://ieeexplore.ieee.org/document/9006456 (regarding the use of AI to distinguish gunfire from other sounds like fireworks); RYAN LILIEN, U.S. DEP'T OF JUSTICE, DEVELOPMENT OF COMPUTATIONAL METHODS FOR THE AUDIO ANALYSIS OF GUNSHOTS 16 (2019), https://www.ojp.gov/pdffiles1/nij/grants/252947.pdf (regarding the potential of AI to assess gunfire and identify the number of shots and types of weapons).

32. Morehead et al., *supra* note 31, at 3044.

33. Faiza Patel, *Advances in AI Increase Risks of Government Social Media Monitoring*, BRENNAN CTR. FOR JUST. (Jan. 4, 2023), https://www.brennancenter.org/our-work/analysis-opinion/advances-ai-increase-risks-government-social-media-monitoring ("[P]olice departments around the country are reviewing and analyzing people's online activity."). College police departments seem particular likely to use social media analysis for threat assessment, as recent news reports indicate. *See* Grace McFadden, *UConn Police Use AI to Track Student Social Media*, THE DAILY CAMPUS (Oct. 3, 2022), https://dailycampus.com/2022/10/03/uconn-police-use-ai-to-track-student-social-media/; Piper Hansen, *ASU Used Software Designed to Monitor Social Media for Safety, Security Threats*, THE ST. PRESS (Sept. 22, 2022, 9:25PM), https://www.statepress.com/article/2022/09/asu-used-social-sentinel-surveillance-tech-2017-to-2020; La Rissa Vasquez, *UC Davis Police Department Using Artificial Intelligence to Monitor Students' Social Media*, THE CAL. AGGIE, https://theaggie.org/2022/11/03/uc-davis-police-department-using-artificial-intelligence-to-monitor-students-social-media/ (last visited Feb. 17, 2024); *see also* Michael Kwet, *Shadowdragon: Inside the Social Media Surveillance Software That Can Watch Your Every Move*, THE INTERCEPT (Sept. 21, 2021, 5:03PM), https://theintercept.com/2021/09/21/surveillance-social-media-police-microsoft-shadowdragon-kaseware/ (discussing the use of social media to create profiles and noting "what used to take us two months in a background check or an investigation is now taking between five to 15 minutes").

34. *See* Vasquez, *supra* note 33.

35. *See* Kwet, *supra* note 33 ("I want to know everything about the suspect: Where do they get their coffee, where do they get their gas, where's their electric bill, who's their mom, who's their dad?").

36. *See* Rigano, *supra* note 29, at 39.

37. *See infra* Section II.B; note 39 and accompanying text (regarding the use of AI technology to create evidence for use at trial).

these technologies in China has led to mass surveillance on a national level, enabling the Chinese government to monitor millions of flagged individuals and causing increased spending on AI research and development.[38] The extent to which using such methods are appropriate in a democratic country like the United States has led to a vigorous debate.[39]

Even if these forensic methods are used by law enforcement, they are less likely to raise direct constitutional challenges.[40] Without courtroom use, the rules of evidence, rules of procedure (criminal or civil), and the Confrontation Clause are irrelevant.[41] To the extent that these technologies are used to identify or charge individual suspects, they do so within the bounds of current Fourth Amendment doctrine.[42] That is not to say that the algorithms are perfect, as they can incorporate biases within their code that exist within the real world, but such claims fall within well-established due process and equal protection doctrines.[43]

As a general matter, the Fourth Amendment protects an individual from unreasonable searches and seizures, and those searches cannot invade a "protected area."[44] Because Supreme Court case law has defined "protected area" mainly based on a reasonable expectation of privacy, the same doctrine will

38. Paul Mozur, *Inside China's Dystopian Dreams: A.I., Shame and Lots of Cameras*, N.Y. TIMES (July 8, 2018), https://www.nytimes.com/2018/07/08/business/china-surveillance-technology.html.

39. *See Know It All: AI and Police Surveillance,* NPR (Feb. 23, 2023, 6:00PM), https://www.npr.org/2023/02/23/1159084476/know-it-all-ai-and-police-surveillance.

40. *See* Srivats Shankar, *Fourth Amendment Constraints on Automated Surveillance Technology in the Public to Safeguard the Right of an Individual to Be "Secure in Their Person*," 18 J. BUS. & TECH. L. 209, 234 (2023) (analyzing the Fourth Amendment implications of different varieties of automated surveillance and explaining that passive searches "implicate a lower privacy expectation [which] is supported by the government's existing practice of identifying an individual through similar technologies").

41. *See Right to Confront Witness*, CORNELL L. SCH.: LEGAL INFO. INST., https://www.law.cornell.edu/wex/right_to_confront_witness (last visited Feb. 2, 2024) (explaining that the Sixth Amendment's Confrontation Clause includes "the right to be present at the trial" and "the right to cross-examine the prosecution's witnesses"); FED. R. EVID. 1101 (explaining the court proceedings in which the Federal Rules of Evidence apply); FED. R. CIV. P. 1 (establishing that the Federal Rules of Civil Procedure govern "civil actions and proceedings in the United States district courts"); FED. R. CRIM. P. 1 (establishing that the Federal Rules of Criminal Procedure "govern the procedure in all criminal proceedings in the United States district courts, the United States courts of appeals, and the Supreme Court of the United States").

42. *See* Shankar, *supra* note 40, at 234 (finding that though use of automated surveillance to identify individuals is common, "using automated surveillance technology to gather more information beyond pure identification may implicate additional concerns under the Fourth Amendment.").

43. *See* Grimm et al., *supra* note 5, at 42 (describing that bias leading to "discriminatory outcomes is a serious problem with AI").

44. Katz v. United States, 389 U.S. 347, 361 (1967) (Harlan, J., concurring).

not protect that which is exposed to the public.[45] The "plain view" doctrine therefore generally prevents a criminal defendant from claiming a privacy interest in anything exposed to plain view.[46] If so, the use of AI technology for, say, examining the faces of individuals in a train station, would only examine that which is exposed to public view and not violate the Fourth Amendment, nor would examining messages posted to social media and viewable by the public. These methods are therefore commonly used for general surveillance, lead generation, or as an accompaniment to a larger investigation, although usually not as evidence in a particular case.[47] In fact, several cities have already made official police policies to that effect. In New York City, for example, facial recognition has explicitly been stated to be used solely as a tool for leads and is not enough for a finding of probable cause.[48] Similarly, the Bureau of Justice Assistance has proposed a standard for law enforcement agencies nationwide explicitly recognizing that facial identification from AI is not enough to establish probable cause and should be subject to additional safeguards.[49] For this reason, facial identification is unlikely to be offered in court.[50]

45. *See, e.g.*, Florida v. Jardines, 569 U.S. 1, 5 (2013) (explaining that the extent to which the *Katz* reasonable expectation of privacy test has been supplanted by analysis of property rights in recent Supreme Court cases remains an open question).

46. Minnesota v. Dickerson, 508 U.S. 366, 374–75 (1993) (reviewing prior precedent addressing when plain view doctrine permits observation and, with lawful access, seizure of an object observable and incriminating to police).

47. *See NYPD Questions and Answers Facial Recognition*, https://www.nyc.gov/site/nypd/about/about-nypd/equipment-tech/facial-recognition.page (last visited Feb. 2, 2024) ("A facial recognition match does not establish probable cause to arrest or obtain a search warrant, but serves as a lead for additional investigative steps.").

48. *See id.*

49. BUREAU OF JUSTICE ASSISTANCE, FACE RECOGNITION: POLICY DEVELOPMENT TEMPLATE 4 (2017), https://bja.ojp.gov/sites/g/files/xyckuh186/files/Publications/Face-Recognition-Policy-Development-Template-508-compliant.pdf ("Face recognition search results are not considered positive identification and do not establish probable cause, without further investigation; rather, they are advisory in nature as an investigative lead only.").

50. *See infra* Section II.B (discussing that if law enforcement were to, for example, base a decision of probable cause solely upon the findings of the forensic AI use, then the problematic aspects of evidentiary rules would limit their consideration). Unfortunately, police use of AI to provide probable cause for arrest is not a hypothetical, as recent reported cases demonstrate. *See e.g.*, Drew Harwell, *Wrongfully Arrested Man Sues Detroit Police over False Facial Recognition Match*, WASH. POST (April 13, 2021, 4:18PM), https://www.washingtonpost.com/technology/2021/04/13/facial-recognition-false-arrest-lawsuit/; Jennifer Valentino-DeVries, *How the Police Use Facial Recognition, and Where It Falls Short*, N.Y. TIMES (Jan. 12, 2020), https://www.nytimes.com/2020/01/12/technology/facial-recognition-police.html. Yet even here, the legal system has remedies in place to address

That is not to say that the government should allow unfettered AI usage either. Use of AI for mass surveillance raises additional privacy concerns beyond the limitations of criminal justice and is appropriately the subject of debate on proper limitations and safeguards.[51] The European Parliament adopted a resolution banning the use of AI facial identification for general, sweeping use, citing concerns about intrusion into private lives, protection of privacy, and the desire to live in a society without governmental surveillance or social scoring.[52] The EU is considering similar limitations in the area of predictive policing as well.[53]

The United States should, for the same reasons, consider adopting legislative or regulatory limits on use of AI in mass surveillance. Representatives in Congress have made proposals to this effect, specifically addressing facial recognition, but so far their efforts have not led to legislation.[54] The debate on AI limits can also be seen at the state level. Several states have adopted legislation that allows individuals the right to opt out of use of their personal data for profiling purposes, in any matter with significant legal, educational, or financial effects.[55] Specific legislation has also been enacted to address particular concerns in several states, such as Illinois (use of AI in hiring), Colorado

---

the concern. *See* Sarah Hughes Newman, *Providing Probable Cause: Allocating the Burden of Proof in False Arrest Claims under § 1983*, 73 U. CHI. L. REV. 347, 347–48 (2006) (explaining that an individual arrested without probable cause can bring a civil action under 42 USC § 1983). An individual subject to arrest on a finding of probable cause through AI would be able to appropriately challenge their detention, to quash the warrant or accusation, and to seek civil damages for wrongful arrest under 42 U.S.C. § 1983 or other similar statutes. *See* KENT BRINTNALL, U.S. COURT OF APPEALS FOR THE NINTH CIRCUIT, SECTION 1983 OUTLINE 47–50 (2011) (explaining potential remedies under § 1983 claims).

51. See Mozur, *supra* note 38; *Know It All: AI and Police Surveillance*, *supra* note 39 (regarding the use of AI for mass surveillance in other countries and the debate on its use in the United States).

52. Lisa Peets et al., *European Parliament Votes in Favor of Banning the Use of Facial Recognition in Law Enforcement*, INSIDE PRIVACY (Oct. 12, 2021), https://www.insideprivacy.com/artificial-intelligence/european-parliament-votes-in-favor-of-banning-the-use-of-facial-recognition-in-law-enforcement/.

53. James Vincent, *EU Draft Legislation Will Ban AI for Mass Biometric Surveillance and Predictive Policing*, THE VERGE (May 11, 2023, 8:19AM), https://www.theverge.com/2023/5/11/23719694/eu-ai-act-draft-approved-prohibitions-surveillance-predictive-policing.

54. *See, e.g.*, Facial Recognition and Biometric Technology Moratorium Act of 2020, S. 4084, 116th Cong. (2020).

55. *See US State-by-State AI Legislation Snapshot*, BRYAN CAVE LEIGHTON PAISNER, https://www.bclplaw.com/print/v2/content/1519741/2023-state-by-state-artificial-intelligence-legislation-snapshot.pdf (last visited Feb. 2, 2024) (providing an overview of the current state of legislation on AI in the United States). States with rules limiting use of personal data for profiling include California, Colorado, and Connecticut. *Id.*

(use of AI for setting insurance rates), and New York (use of AI for employment decisions).[56] Nearly half of the states are considering additional legislation to regulate AI in a variety of fields.[57]

Beyond regulatory or legislative proposals, the lack of information on the prevalence of AI usage in law enforcement should make the need for data and openness clear.[58] Therefore, whether governmental actors provide details or information, citizens themselves can request information on adoption and usage by state and federal sunshine laws such as the Freedom of Information Act.[59] Should such efforts lead to disclosure of additional abuses in AI use in mass surveillance, then certainly additional regulatory or legislative efforts could be necessary. But without the raw data, accountability will be difficult to obtain.

The extent to which AI should be exposed, regulated, and assessed, plus the potential for it to contain improper biases and the proper reaction to that potential, will be explored in a later article.

B. Use of Artificial Intelligence by Law Enforcement—Forensic Evidentiary Use

Use of AI for lead generation or mass surveillance may involve one set of considerations, largely demanding regulatory or legislative oversight.[60] But unlike facial recognition, social media surveillance, or predictive policing, law enforcement can and does use AI to assess evidence and provide conclusions related to an individual case.[61] As with mass surveillance, the type, extent,

56. *Id.*

57. *See id.*

58. *See* Kwet, *supra* note 33. Some of the information regarding AI usage in modern policing has been obtained by precisely this method. *See id.* ("What's more, Goldberg had to file a Freedom of Information Act request to obtain the contract.").

59. Freedom of Information Act, 5 U.S.C. § 552 (2018); *see also State Freedom of Information Laws*, NAT'L FREEDOM INFO. COALITION, https://www.nfoic.org/state-freedom-of-information-laws/ (last visited Feb. 19, 2024) (providing a compilation of similar state statutes).

60. *See* KRISTIN FINKLEA, CONG. RESEARCH SERV., IN12289, LAW ENFORCEMENT USE OF ARTIFICIAL INTELLIGENCE AND DIRECTIVES IN THE 2023 EXECUTIVE ORDER 2 (2023), https://crsreports.congress.gov/product/pdf/IN/IN12289#:~:text=AI%20can%20be%20used%20along,of%20being%20involved%20in%20crime.&text=Law%20enforcement%20agencies%20can%20employ,and%20push%20out%20emergency%20information ("Policymakers may consider increased oversight over police use of AI systems to help evaluate and alleviate some of the shortcomings.").

61. *See* Chris Hsiung & Frank Chen, *Exploring AI for Law Enforcement*, POLICE CHIEF (Sept. 20, 2023), https://www.policechiefmagazine.org/exploring-ai-law-enforcement-interview/ ("AI systems will help solve crimes by making it easier to gather, analyze, and act on evidence.").

and prevalence of evidence-generative AI in law enforcement is not known except inferentially through public records, media reports, or judicial opinions.[62] To the extent forensic AI is and will be used for evidence at trial, however, it raises a much different set of concerns which relate to courtroom rules and evidentiary limits central to this Article.

Several examples of forensic AI will help illustrate the type of evidence which can and will be offered in courtrooms today or in the near future. These examples are not meant to be an exhaustive list of all usages but provide basic insights into the type of forensic fields where AI has potential to be used. The examples show that the use of forensic AI is already wide-ranging.

Nowhere is AI deployment for forensic use more advanced than in the area of DNA analysis. Samples obtained for forensic analysis can contain multiple contributors, thus interpretating these samples and determining a match is exceedingly difficult.[63] For that reason, research into computer-aided analysis of DNA samples has been extensive in the past decade.[64] Several of these methods are computational in design and, for that reason, stand outside the definition of "artificial intelligence."[65] Other programs, most notably TrueAllele, but likely others in the near future, do use AI to analyze the

62. *See* Grimm et al., *supra* note 5, at 105 (explaining that although regulation regarding the admissibility of AI as evidence is not well-developed, some courts have addressed the admissibility of AI evidence in proceedings governed by rules of evidence).

63. FAIGMAN ET AL., *supra* note 8, § 30:34 (stating that for some cases, the analysis may be "insurmountably difficult").

64. *See, e.g.*, *A Fully Continuous Machine Learning Approach to Predict the Number of Contributors in Sequence-Based DNA Profiles*, NAT'L INST. JUSTICE (Sept. 27, 2018), https://nij.ojp.gov/funding/awards/2018-du-bx-0202; *A Hybrid Machine Learning Approach for DNA Mixture Interpretation*, NAT'L INST. JUSTICE (Sept. 11, 2014), https://nij.ojp.gov/funding/awards/2014-dn-bx-k029?award=2014-DN-BX-K029. The 2014 research grant alone led to several published studies in the field, such as Michael Marciano et al., *A Hybrid Approach to Increase the Informedness of CE-Based Data Using Locus-Specific Thresholding and Machine Learning*, 35 FORENSIC SCI. INT'L: GENETICS 26, 26 (2018); Michael A. Marciano et al., *PACE: Probabilistic Assessment for Contributor Estimation—A Machine Learning-Based Assessment of the Number of Contributors in DNA Mixtures*, 27 FORENSIC SCI. INT'L: GENETICS 82, 82 (2017); Michael A. Marciano & Kevin S. Sweder, *Hybrid Machine Learning Approach for DNA Mixture Interpretation*, NAT'L INST. JUSTICE (June 1, 2016), https://nij.ojp.gov/library/publications/hybrid-machine-learning-approach-dna-mixture-interpretation.

65. People v. H.K., 130 N.Y.S.3d 890, 898 (N.Y. Crim. Ct. 2020) (noting that unlike TrueAllele, STRMix does not rely on AI); *see also* Letter from John Buckleton, D.Sc., to Justice Jill Presser, Superior Court of Ontario (Aug. 9, 2021), https://johnbuckleton.files.wordpress.com/2021/08/ai-case-study-ii.pdf. The computational method that STRMix uses to assess likelihood ratios is described in detail in *People v. Davis*, 290 Cal. Rptr. 3d 661, 677–79 (Ct. App. 2022). *See also* FAIGMAN ET AL., *supra* note 8, § 30:34 (describing STRMix, TrueAllele, and other mixture analysis software in detail).

genotypes submitted.[66] A forensic examiner using TrueAllele, then, would receive a likelihood ratio from a sample, indicating "how much more probable the observed result would be . . . if the contributor was the suspect than if the contributor was a random person."[67] Such DNA evidence at trial can provide powerful evidence of guilt against the accused,[68] particularly considering DNA remains the model of rigor in forensics by the National Research Council and has been described as "one of the most significant scientific advancements of our era" by the Supreme Court.[69]

DNA is not the sole identification method that can incorporate AI technology. AI may be used for a variety of fingerprint analysis tasks, specifically for image acquisition, image enhancement, feature extraction, or even, potentially, for matching.[70] Researchers at West Virginia University have used AI to enhance fingerprint images for analysis, allowing over 95% of their initial blurred images to be matched once deblurred by AI.[71] A separate company markets an AI tool that enhances detection and then details of fingerprints for analysis by examiners, promising "unerring assistance to the human examiner."[72] Medical examiners see AI's potential for fingerprint analysis as well—some researchers anticipate a fully autonomous AI fingerprint-

66. People v. Wakefield, 175 A.D.3d 158, 162 (N.Y. App. Div. 2019) (noting the founder and chief scientist at Cybergenetics, the company that sells TrueAllele, testified that it does contain AI); *H.K.*, 130 N.Y.S.3d at 898 (noting that unlike TrueAllele, STRMix does not rely on AI); *see* Letter from John Buckleton, *supra* note 65.

67. William C. Thompson et al., *Forensic DNA Statistics: Still Controversial in Some Cases*, THE CHAMPION, Dec. 2012, at 12, 19; *see* FAIGMAN ET AL., *supra* note 8, § 30:27 (defining the likelihood ratio for DNA analysis); *see also* Roth, *supra* note 18, at 1262.

68. The effect is so powerful that it has been reduced to a shorthand: If DNA then Guilty. *See* Christina Kline, et al., '*If DNA, Then Guilty': Strategies for Overcoming Juror Assumptions About DNA Evidence in Criminal Trials*, THE CHAMPION, Jan.–Feb. 2015, at 22.

69. NRC REPORT, *supra* note 9, at 7; *see also* Maryland v. King, 569 U.S. 435, 442 (2013) ("The advent of DNA technology is one of the most significant scientific advancements of our era. The full potential for use of genetic markers in medicine and science is still being explored, but the utility of DNA identification in the criminal justice system is already undisputed.").

70. Kenneth R. Moses et al., *Automated Fingerprint Identification System*, *in* THE FINGERPRINT SOURCEBOOK 6-1, 6-20 to 6-27 (2011) (stating that better matching using AI is only a potential, as current systems are less accurate than human forensic experts).

71. Amol S. Joshi et al., *FDeblur-GAN: Fingerprint Deblurring Using Generative Adversarial Network*, W. VA. U. (June 21, 2021), https://arxiv.org/abs/2106.11354.

72. *Meet AARI, the Unique New Fingerprint Imaging System That Uses Artificial Intelligence to Detect Ridge Detail and Assist Forensic Examiners*, FOSTER+FREEMAN, https://www.google.com/url?q=https://fosterfreeman.com/meet-aari-the-unique-new-fingerprint-imaging-system-that-uses-artificial-intelligence-to-detect-ridge-detail-and-assist-forensic-examiners/&sa=D&source=editors&ust=1708472905840760&usg=AOvVaw2OTGMsAZ-TMO3PzVkoqcuzS (last visited Feb. 20, 2024).

matching tool to be able to identify unknown decedents using fingerprint matches to prior samples in a database.[73]

In addition to the identification purposes, forensic medical analysis could also be an area where AI tools are employed. In one research project, doctors developed an AI tool to take medical imaging of infants with head trauma and calculate the characteristics of force that would lead to an injury of that magnitude, such as fall height, angle of impact, and impact location.[74] Similar studies have been funded by the National Institute of Justice and remain in the research phase, including a study to quantify the assessment of bruise age in forensic trauma analysis,[75] and another to enhance accuracy of ancestry estimation from cranial analysis.[76] Researchers also anticipate AI usage in other medical assignments such as determinations of time of death, toxicological analysis, or autopsy assistance through organ pathology analysis.[77]

Voice identification AI could also serve forensic purposes for use in court. A recent study in *Forensic Science International* reported that an AI tool for forensic voice comparison outperformed human listeners in voice matching and identification.[78] Even more impressively, the research used conditions

73. Toshal D. Wankhade et al., *Artificial Intelligence in Forensic Medicine and Toxicology: The Future of Forensic Medicine*, CUREUS (Aug. 25, 2022), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9506671/ ("When the particular pattern of biometry of the individual, e.g[.], a fingerprint, is brought in front of the machine, it will quickly recognize that individual by matching the biometric pattern which is already stored in the machine. Thus, individual identity can be established by matching the biometric pattern put in front of a machine with the biometric data of the individual already stored at the machine.").

74. Tagrid M. Ruiz-Maldonado et al., *Age-Related Skull Fracture Patterns in Infants After Low-Height Falls*, 93 PEDIATRIC RES. 1990, 1990 (2022). A description of the method of analysis appears in the grant proposal at the National Institute for Justice. *Forensic Tool to Identify Fall Characteristics in Infant Skull Fracture,* NAT'L INST. JUSTICE, https://nij.ojp.gov/funding/awards/2020-75-cx-0014 (last visited Feb. 20, 2024) (identifying the specific methodology as using a machine learning algorithm).

75. *Collaborative Artificial Intelligence Platform for Bruise Age Analysis*, NAT'L INST. JUSTICE, https://nij.ojp.gov/funding/awards/15pnij-21-gg-04145-slfo (last visited Feb. 20, 2024).

76. *Improve Craniometric Ancestry Estimation with Deep Learning Methods*, NAT'L INST. JUSTICE, https://nij.ojp.gov/funding/awards/15pnij-22-gg-04431-ress (last visited Feb. 20, 2024).

77. *See generally* Wankhade et al., *supra* note 73 (describing how AI technology can be helpful for forensic medical analysis where it might "play a key role in aiding forensic experts to form more accurate, quick, and uniform opinions related to forensic case examination by comparing the data from their findings with the data available from machines").

78. Nabanita Basu et al., *Speaker Identification in Courtroom Contexts—Part I: Individual Listeners Compared to Forensic Voice Comparison Based on Automatic-Speaker-Recognition Technology*, 341 FORENSIC SCI. INT'L 1, 20 (2022), https://www.sciencedirect.com/science/article/pii/S0379073822003292.

similar to real-world forensic casework.[79] In similar work, the Speaker Identification Integrated Project in Europe has been a coordinated effort to create and deploy a speaker identification system at Interpol.[80] Specifically, the researchers intend "to develop 'a system that identifies voices in audio sourced from lawfully intercepted communication and social media.'"[81] Beyond use by public authorities, AI-enhanced voice ID has been and will be deployed in the private sector for usages such as identifying confirmation, banking security, or even healthcare or marketing.[82]

Finally, AI might lead to enhanced use of lie detection. Traditional lie detectors, initially offered in the first half of the twentieth century, use biometric data (heart rate, blood pressure, respiration) to attempt measuring the truth or falsity of a spoken statement.[83] The results of polygraph examination have not traditionally been admissible in court, whether due to lack of reliability of the techniques or a perception that they infringe on the role of the jury.[84] Researchers at several universities in the United States have taken AI tools to analyze written text, finding they could identify falsity at a rate sometimes exceeding 80%.[85] Assessment of spoken word deception using AI has also been studied extensively, with promising results that suggest the potential for future forensic deployment. A 2022 RAND Corporation study found spoken work assessment of deception could be as high as 76%, depending on the

79. *Id.* at 7–8.

80. Fieke Jansen et al., *Biometric Identity Systems in Law Enforcement and the Politics of (Voice) Recognition: The Case of SiiP*, 8 BIG DATA & SOC'Y 1, 1 (2021), https://journals.sagepub.com/doi/epub/10.1177/20539517211063604. *See generally Speaker Identification Integrated Project,* CORDIS, https://cordis.europa.eu/project/id/607784 (last visited Feb. 2, 2024).

81. Jansen et al., *supra* note 80, at 5.

82. *How Does Voice Recognition Biometrics Work?*, NEC N.Z. (Jan. 26, 2022), https://www.nec.co.nz/market-leadership/publications-media/how-does-voice-recognition-biometrics-work/.

83. Katherine To, *Lie Detection: The Science and Development of the Polygraph*, ILLUMIN MAG. (Dec. 6, 2002), https://illumin.usc.edu/lie-detection-the-science-and-development-of-the-polygraph/ (describing the polygraph test's development and reliance on physiological changes to detect dishonesty).

84. *See* FAIGMAN, *supra* note 8, § 38:3 (noting the exclusion of polygraph tests in many courts in the United States, as well as the rationale for the exclusions).

85. ALBERTO ALEJANDRO CEBALLOS DELGADO ET AL., DETECTING DECEPTION USING MACHINE LEARNING, PROCEEDINGS OF THE 54TH HAWAII INTERNATIONAL CONFERENCE ON SYSTEM SCIENCES 7128 (2021), https://shsu-ir.tdl.org/server/api/core/bitstreams/3a674707-3f54-476a-aab9-17bc7698efa1/content.

model used for analysis.[86] Researchers have used AI to assess deception using analysis of facial expressions, leading to deception detection rates also in the 70s.[87] Needless to say, these usages also raise significant ethical issues.[88]

While these examples are not intended to be an exhaustive list of all usages, they provide basic insights into the current use and likely future developments in the field. What they all have in common is that in these usages, the AI technology is being deployed to reach a conclusion that is often solely the purview of an expert witness.[89] This is true whether the AI is telling communicating the likelihood ratio of a match in a DNA mixture, a voice comparison between a known and unknown sample, the likelihood of organ damage as a contributing cause of death, the timing of an injury to the victim, or the truthfulness of a witness (although this has, to date, remained a question solely for the jury).[90]

The big question, then, is whether and when each type of AI forensic expert conclusion is admissible in court as evidence, and this question is likely to be of greater significance as these technologies advance and deploy in real world situations.[91]

To understand how current doctrine applies to evidence generation AI technology, it is necessary to first explain what they have in common—the core function and methods of AI itself. Only then can we apply current doctrine to determine the limits of admissibility now and whether further developments in the law are appropriate.

III. COMPUTATIONAL METHODOLOGIES OF AI

On January 1, 2021, Congress enacted the National Artificial Intelligence

86. Marek N. Posard et al., *Deception Detection*, RAND (Oct. 19, 2022), https://www.rand.org/pubs/research_briefs/RBA873-1.html.

87. Merylin Monaro et al., *Detecting Deception Through Facial Expressions in a Dataset of Videotaped Interviews: A Comparison Between Human Judges and Machine Learning Models*, 127 COMPUTERS HUM. BEHAV. 1, 2 (2022); Anastasia Shuster et al., *Lie To My Face: An Electromyography Approach to the Study of Deceptive Behavior*, 11 BRAIN & BEHAV. 1, 2 (2021).

88. Jo Ann Oravec, *The Emergence of "Truth Machines"?: Artificial Intelligence Approaches to Lie Detection*, 24 ETHICS & INFO. TECH. 1, 6 (2022).

89. *See* Grimm et. al., *supra* note 5, at 79–82 (discussing the ambiguities that arise when prosecutors utilize AI technology to achieve the information and cognition of an expert witness).

90. *See* Roth, *supra* note 5, at 1978–79 (emphasizing that despite humans having control over what is input into machines, machines have credibility issues of their own).

91. *Id.* at 1975–76 (highlighting the courtroom's shift towards machine evidence over the past century).

Initiative with the following purposes:

> (1) ensure continued United States leadership in artificial intelligence research and development;
>
> (2) lead the world in the development and use of trustworthy artificial intelligence systems in the public and private sectors;
>
> (3) prepare the present and future United States workforce for the integration of artificial intelligence systems across all sectors of the economy and society; and
>
> (4) coordinate ongoing artificial intelligence research, development, and demonstration activities among the civilian agencies, the Department of Defense and the Intelligence Community to ensure that each informs the work of the others.[92]

Under the National Artificial Intelligence Initiative, "artificial intelligence" is defined as "a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations or decisions influencing real or virtual environments."[93]

As the Initiative makes clear, there are multiple ways that machines can engage in these acts of AI.[94] In this section, we will first discuss the central methodological problems with using artificially intelligent experts: their lack of explainability and transparency, and the related concepts of open-box and closed-box systems. We then discuss, in turn, three categories of AI systems: expert systems, machine learning, and neural networks. As we discuss each system, we will explain if the problems of transparency and explainability can be addressed in the system and to what degree, as well as identifying whether and when the system is open-box or closed-box.

A. Explainability and Transparency: The Central Issue for AI Experts

AI developers encounter three basic challenges. First, they must worry

92. 15 U.S.C.A. § 9411(a) (West 2021).
93. 15 U.S.C.A. § 9401(3) (West 2021).
94. 15 U.S.C.A. § 9411(b) (West 2021).

about the energy and computational costs of the system.[95] For example, deep learning models are extremely complex and energy-expensive. As a result, AI is limited in application to devices capable of supporting its architecture and energy use.[96] Second, if the system is out in the world, it is subject to attack by external actors (hacking).[97] For example, the "[a]utomobile manufacturer Chrysler announced a recall of 1.4 million vehicles when a pair of hackers, Charlie Miller and Chris Valasek, demonstrated their ability to remotely hijack the digital systems of a Jeep through the Internet."[98] In particular, the researchers were "able to send commands to the engine and wheels."[99] Finally, AI systems are subject to the problem of understanding why they come to the decisions that they do—these are the problems of transparency and explainability.[100] We will consider AI systems that are transparent and explainable to be open-box systems and those AI systems that are neither transparent nor explainable to be closed-box systems. In general, we find that expert systems are open-box systems while machine learning systems and neural networks are closed-box systems.

Because of the Sixth Amendment's Confrontation Clause and the Fourteenth Amendment's Due Process Clause, the central problems for AI experts are problems of transparency and explainability. The Sixth Amendment's Confrontation Clause provides that "[i]n all criminal prosecutions, the accused shall enjoy the right . . . to be confronted with the witnesses against him."[101] Confrontation means the right to cross-examine, to ask the witness to explain oneself.[102] Similarly, in the civil context, due process requires that "[a]ll parties must be fully apprised of the evidence submitted or to be considered, and must be given opportunity to cross-examine witnesses, to inspect documents, and to offer evidence in explanation or rebuttal."[103]

95. Wojciech Samek & Klaus-Robert Müller, *Towards Explainable Artificial Intelligence*, *in* EXPLAINABLE AI: INTERPRETING, EXPLAINING AND VISUALIZING DEEP LEARNING 5, 6 (Wojciech Samek et al. eds., 2019) [hereinafter EXPLAINABLE AI].

96. *See, e.g.*, ETHEM ALPAYDIN, INTRODUCTION TO MACHINE LEARNING 215–26 (Thomas Dietterich ed., 3d. ed. 2014).

97. Samek & Müller, *supra* note 95, at 6.

98. Alexandra Green, *The Self Drive Act: An Opportunity to Re-Legislate a Minimum Cybersecurity Federal Framework for Autonomous Vehicles*, 60 SANTA CLARA L. REV. 217, 229 (2020).

99. *Id.* at 230.

100. Samek & Müller, *supra* note 95, at 6.

101. U.S. CONST. amend. VI.

102. *See, e.g.*, Crawford v. Washington, 541 U.S. 36, 45–47 (2004).

103. Interstate Commerce Comm'n v. Louisville & Nashville R.R. Co., 227 U.S. 88, 93 (1913); *see also* Goldberg v. Kelly, 397 U.S. 254, 269–70 (1970).

As we will see, the issue of transparency falls within the domain of the attorney.[104] In essence, the party's legal representative must have sufficient access to the system so that their experts can interpret them.[105] On the other hand, the issue of explainability will fall within the domain of the fact finder who will need to understand why the AI system produced the result it did without needing to be able to look directly under the hood of the AI system.[106]

B. Right Answers for the Wrong Reasons

The point of confrontation is to enable the fact finder to determine the truth of the matter testified to.[107] Cross-examination

> (1) insures that the witness will give his statements under oath—thus impressing him with the seriousness of the matter and guarding against the lie by the possibility of a penalty for perjury;
>
> (2) forces the witness to submit to cross- examination, the "greatest legal engine ever invented for the discovery of truth";
>
> (3) permits the jury that is to decide the defendant's fate to observe the demeanor of the witness in making his statement, thus aiding the jury in assessing his credibility.[108]

Furthermore, the inspection of documents also aids in the discovery of truth.[109]

The question we must ask of the rules of evidence is whether they

104. *See* Grimm, *supra* note 5, at 105 (recommending attorneys bring all AI evidence to the court's attention as early as possible to assess the evidence's credibility).

105. *Id.* at 97–101.

106. *See* Patrick W. Nutter, *Machine Learning Evidence: Admissibility and Weight*, 21 U. PA. J. CONST. L. 919, 950 (2019) (explaining that a jury's understanding of AI evidence may affect its decision).

107. *See, e.g.*, Dutton v. Evans, 400 U.S. 74, 89 (1970) ("[T]he mission of the Confrontation Clause is to advance a practical concern for the accuracy of the truth-determining process in criminal trials by assuring that 'the trier of fact (has) a satisfactory basis for evaluating the truth of the prior statement.'" (quoting California v. Green, 399 U.S. 149, 161 (1970))).

108. *Green*, 399 U.S. at 158 (internal citation omitted).

109. *See, e.g.*, United States v. Noble, 422 U.S. 225, 232 (1975); Hilton v. Guyot, 159 U.S. 139, 142 (1895); Welzel v. Bernstein, 233 F.R.D. 185, 186 (D.D.C. 2005); Belcher v. Bassett Furniture Indus., Inc., 588 F.2d 904, 908 (4th Cir. 1978).

improve the ability of the fact finder to identify the truth.[110] Unlike in the case of a human expert, a nonhuman expert will take no oath, have no demeanor to observe, nor feel any pressure to tell the truth out of a fear of a penalty of perjury or because of an understanding of the seriousness of its testimony.[111] This requires us to invent new modes of ensuring truthfulness of the evidence offered. To motivate our concerns, this subsection now describes two scenarios where nonhuman intelligences were able to provide the correct answers to queries made to them, but they did so for the wrong reasons.

Starting in 1891, a horse by the name of Clever Hans became a scientific sensation when it was found that he "could apparently perform mathematical calculations, tell time, identify musical intervals, and name people."[112] Relative to mathematical calculations, Hans was able to correctly answer about 90% of the time by tapping out the answer.[113] From this, scientists and Hans's owner, William von Osten, believed that Hans had a high level of intelligence.[114] Sixteen years after the world was introduced to Clever Hans, "a group of thirteen scientists (the 'Hans Commission') re-tested Clever Hans" using a very carefully controlled psychological experiment.[115] As a result, the scientists learned that Clever Hans could not actually do math but rather, he was extremely good at reading the questioner's subtle body language to know when to stop tapping.[116] This experiment made Clever Hans's methodology (intelligence) transparent and provided an explanation for his (usually) correct answers.[117] It also showed that Clever Hans's right answers were not justified by the "right" reasons.[118]

110. *See* Gen. Elec. Co. v. Joiner, 522 U.S. 136, 150 (1997) (Breyer, J., concurring).

111. *See, e.g.*, United States v. Gonzalez, 559 F.2d 1271, 1273 (5th Cir. 1977) (noting the importance of trustworthiness in hearsay statements where a declarant is unavailable); State v. Spadafore, 220 S.E.2d 655, 664 (W. Va. 1975) (holding that "while prior statements made under oath in a judicial atmosphere either by deposition or at a prior trial and which have been subject to cross-examination by the defendant's counsel are admissible for the truth of the matter asserted, all other out-of-court statements may be used exclusively to impeach credibility and should be used sparingly in that regard."); Spigarolo v. Meachum, 934 F.2d 19, 24 (2d Cir. 1991) ("When children testify, the trial court may fashion an oath or affirmation that is meaningfull to the witness . . . permit[ing] witnesses to declare that they will testify truthfully.").

112. Tegoseak v. State, 221 P.3d 345, 351 n.7 (Alaska Ct. App. 2009).

113. *Id.*

114. *Id.*

115. *Id.*

116. *Id.*

117. *Id.*

118. *Id.* (describing Clever Hans's reliance on the questioner's physical cues to know when to stop tapping its foot at the correct time).

Similarly, a visual identification system was fed a series of photos and asked to develop a methodology for identifying horses.[119] It did so, but after close examination of why, it was discovered that the system was honing in on a copyright tag that appeared only in the bottom-left corner of horse pictures.[120] Another system was trained using pictorial data that enabled the system to develop a reliable methodology for distinguishing huskies from wolves.[121] The system did so not based on any differences between the animals but rather due to the presence of snow in the picture (husky pictures tended to have snow in the background while wolf pictures did not).[122] Thus, our artificial intelligences, like animal intelligences, are capable of getting it right, while also getting it completely wrong.

C. Transparency and Explainability

The Clever Hans, husky/wolves, and horse identification examples show us that an AI system can reliably produce the correct results, even if it is following a methodology that does not accurately represent the correct rationale for such a result. This exposes the central weaknesses of an AI expert in our search for truth: how do we get it to explain its answers, and how do we investigate whether what it says it does is correct? These are the dual problems of explainability and transparency. And our rules of evidence will need to determine precisely when an AI system is sufficiently explainable and transparent.

Roughly, a system is explainable if, when answering queries, it presents information to the user that provides a qualitative understanding of the connection between the input and the output.[123] We contend that, relative to the law of evidence, explainability has three key requirements: (1) Fidelity—the explanation must reasonably represent what the system did to produce the output; (2) Understandability—the explanation must be understandable to the person receiving it; and (3) Sufficiency—the explanation must be sufficiently

119. Sebastian Lapuschkin, Opening the Machine Learning Black Box with Layer-wise Relevance Propagation 70–71 (2019) (Ph.D. thesis, Technische Universität Berlin).

120. *Id.*

121. Marco Tulio Ribeiro et al., *Why Should I Trust You? Explaining the Predictions of Any Classifier*, *in* PROCEEDINGS OF THE 22ND ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING 1135, 1142–43 (2016).

122. *Id.* at 1142.

123. *See id.* at 1136.

detailed to justify its output relative to this inquiry.[124]

Transparency requires a system's algorithms, data, and models to be sufficiently open and accessible to review.[125] In the context of the law of evidence, we contend that transparency has three key requirements: (1) Accessibility—whether the system has provided sufficient access to, and information about, its algorithms, models, data sources, and decision-making processes;[126] (2) Understandability—whether the system produces output that is easily understood and interpreted by users;[127] and (3) Data Provenance—whether the system provides information as to the origins and processing of the data, and by whom, used by the system.[128]

Systems that are neither transparent nor explainable are closed-box systems.[129] On the other hand, systems that are fully transparent and explainable are open-box systems.[130] Closed-box AI experts clearly fail to meet the confrontation and due process requirements of the Constitution.[131] Whether an open-box AI system meets those challenges will depend on how transparent and explainable the system is.

D. Types of AI Systems

To understand the difficulty of using AI systems as evidence, we must first understand how they work. In this section, we explain the intricacies of two types of AI systems: expert systems and machine learning systems (including neural networks). As we shall see, the various types of expert and machine learning systems are intrinsically different regarding how transparent

124. Lars Kai Hansen & Laura Rieger, *Interpretability in Intelligent Systems—A New Concept?*, *in* EXPLAINABLE AI, *supra* note 95, at 41, 43.

125. *See* Adrian Weller, *Transparency: Motivations and Challenges*, *in* EXPLAINABLE AI, *supra* note 95, at 23, 23–27.

126. *See, e.g.*, Miriam C. Buiten, *Towards Intelligent Regulation of Artificial Intelligence*, 10 EUR. J. RISK REG. 41, 54 (2019).

127. *See* Mara Graziani et al., *A Global Taxonomy of Interpretable AI: Unifying the Terminology for the Technical and Social Sciences*, 56 ARTIFICIAL INTELLIGENCE REV. 3473, 3480 (2023).

128. *See* Weller, *supra* note 125, at 24.

129. *See, e.g.*, Sarah Kamensky, *Artificial Intelligence and Technology in Health Care: Overview and Possible Legal Implications*, 21 DEPAUL J. HEALTH CARE L. 1, 3 (2020).

130. *See, e.g.*, Ashley Deeks, *Will Cyber Autonomy Undercut Democratic Accountability*, 96 INT'L L. STUD. 464, 489 (2020).

131. *See* Brian Sites, *The Future of the Confrontation Clause: Semiautonomous and Autonomous Machine Witnesses*, 22 VAND. J. ENT. & TECH. L. 547, 574 (2020) (comparing the "black box" dangers of machine sources to hearsay dangers in witness testimony, implicating the Confrontation Clause).

and explainable they are, and whether they are open-box or closed-box systems.

1. Expert Systems

An expert system is composed of a database (the "knowledge base") containing relevant information about the subject matter of the system and the program (the "inference engine") that applies the rules of logic and probability to the knowledge base to draw conclusions.[132] We will discuss four categories of expert systems: rule-based systems, case-based systems, Bayesian networks, and fuzzy systems. Rule and case-based expert systems will prove to be the most transparent and explainable AI systems because they act in a way that is very familiar to most people.[133] At the same time, the more sophisticated expert systems relying on statistics and fuzzy logic are less transparent and explainable.[134] Each of these four types of expert systems is open-box, but for some systems, like those based on Bayesian statistics and fuzzy logic, it may be difficult for jurors to fully understand how the system works due to the mathematical complexity of such systems.[135]

Before we examine each type of expert system, it is useful to provide a diagram of its common architecture.

132. Vehicle Intelligence & Safety LLC v. Mercedes-Benz USA, LLC, 78 F. Supp. 3d 884, 889 (N.D. Ill. 2015).

133. *See* Kevin D. Ashley, *Case-Based Reasoning and Its Implications for Legal Expert Systems*, 1 ARTIFICIAL INTELLIGENCE & L. 113, 114 (1992) (explaining case-based systems are designed to assist human reasoning); Anshul Jain, *Rule-Based Systems*, PROFESSIONALAI.COM (Apr. 22, 2020), https://www.professional-ai.com/rule-based-systems.html (explaining that rule-based systems apply knowledge through facts, rules, and determining a course of action).

134. *See* Trung T. Pham & Guanrong Chen, *Some Applications of Fuzzy Logic in Rule-Based Systems*, 19 EXPERT SYS. 208, 208 (2002) (explaining that fuzzy logic allows more contradicting rules to coexist before a system generates the final action).

135. *See* William C. Thompson, *Are Juries Competent to Evaluate Statistical Evidence?*, 52 LAW & CONTEMP. PROBS. 9, 25 (1989) (highlighting the concern that jurors will be disproportionately influenced by complicated statistical evidence).

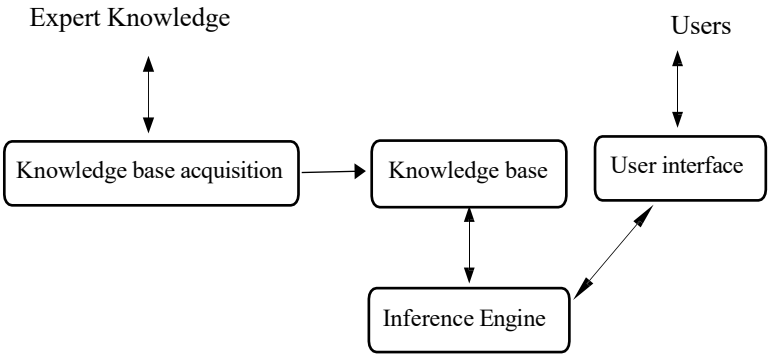

Figure A. Architecture of a simple expert system.[136]

Shown by Figure A, "knowledge base acquisition" is the method of collecting knowledge from the expert; the "knowledge base" is the final product of that knowledge base acquisition process; the "user interface" is the modality through which the user makes queries and gets answers; and the "inference engine" is the system that provides answers to the queries from the user by applying a set of inference rules to the knowledge base.[137]

a. Rule-Based Systems

In a rule-based expert system, the knowledge base contains both facts relevant to the subject matter of the system and "heuristics or rules that control the use of knowledge to solve problems in a particular domain."[138] These rules typically come in the conditional form: "if . . . then . . ." expressions.[139] In a sentence of the form of "if *P*, then *Q*," we call *P* the "antecedent" and *Q* the "consequent."[140] The inference engine then utilizes its search strategy to "determine[] when rules are needed, which rules to select, and how rules

136. *See* Ajith Abraham, *Rule-Based Expert Systems*, *in* HANDBOOK OF MEASURING SYSTEM DESIGN 909, 910 (Peter H. Sydenham & Richard Thorn eds., 2005).

137. *See id.* at 910–11.

138. *In re* Lockwood, 679 F. App'x. 1021, 1028 (Fed. Cir. 2017) (quoting the joint appendix of *In re Lockwood*) (emphasis omitted).

139. *See* Synopsys, Inc. v. Ricoh Co. Ltd., Nos. C 03–2289 MJJ, C 03–4669, 2005 WL 6217119, at *11 (N.D. Cal. Apr. 7, 2005).

140. *See* MERRIE BERGMANN ET AL., THE LOGIC BOOK 20 (Sarah Jaeger ed., 6th ed. 2014).

should be processed."[141] The inference engine then processes the rules by using logical reasoning and the user's query to derive a response.[142]

For example, if we want to build an expert system for determining if it has rained last night, we might have the following knowledge base:

1. If the sidewalk is wet and the sprinklers have not run, then it has rained.
2. If the sidewalk is wet and the sprinklers have run, then it has not rained.

If the user then queries by asking what happens if the sidewalk is wet and the sprinklers have not run, the inference engine will search the rules for those that are applicable to that input, set the values associated with the query to "true," and apply the rules of logic to derive a conclusion.[143] In this case, that it has rained.

This example uses a forward-chaining inference system where the inference engine uses the initial facts given to it to iteratively draw conclusions.[144] It does so by first identifying rules for which it has values for the antecedents of those rules and then concluding that the respective consequents are true.[145] The inference engine then takes any conclusions found to be true as a result of this first pass and feeds them back through the system in the same manner as the initial query.[146] This process continues until the final conclusion is drawn.[147]

In backward-chaining systems, we start with the goal—the hypothesis or conclusion we want to establish.[148] The backward-chaining system then reviews the rules in the database to identify those rules where the consequent is either that conclusion or its negation.[149] The system them examines those rules one at a time to see if it has any data that matches the rule's antecedent.[150]

---

141. KEN PEDERSEN, EXPERT SYSTEMS PROGRAMMING: PRACTICAL TECHNIQUES FOR RULE-BASED SYSTEMS 55 (1989).

142. *See In re Lockwood*, 679 F. App'x. at 1028 (explaining that the task of the inference engine "is to monitor the facts in the data base and execute the action part of those rules that have their situation part satisfied").

143. *See* PEDERSON, *supra* note 141, at 69 (illustrating the process behind this example in Figure 4.8).

144. *See id.* at 55.

145. *See id.* at 73.

146. *See id.* at 73–74.

147. *See generally id.* at 73–80 (walking the reader through the different steps of the forward-chaining inference system, including how the process reaches its conclusion).

148. *See e.g.*, *id.* at 57 (illustrating this idea using a hypothetical).

149. *Id.*

150. *Id.*

If it does, then the process stops and the conclusion is drawn.[151] If the system does not have sufficient information to determine if the antecedent is true or false, then it sets it as a goal and the process starts over with the system looking for a rule that has the new goal (the antecedent from the original rule) as the consequent.[152] If at the end of that process, the antecedent cannot be determined to be true or false, then the antecedent is deemed by the system to be false.[153] This process continues until either a conclusion is drawn or the system identifies relevant information it needs, but does not have, to draw a conclusion.[154]

One important distinction between forward- and backward-chaining inference engines is that backward-chaining systems start with a goal while forward-chaining systems start with data.[155] Because goals are more subjective than data, this can make backward-chaining inference engines more susceptible to bias than forward-chaining systems.[156]

Rule-based expert systems should easily meet the requirements of being explainable and transparent. Given the simplicity of the system, fact finders are likely to grasp the reasoning process used by the inference engine. In addition, it should be relatively easy to provide explanations of the output that meet the requirements of fidelity, understandability, and sufficiency. Similarly, such systems should easily meet the requirements of transparency—accessibility, understandability, and data provenance.

At the same time, because the system or its data may be proprietary, the developers of the system may want to water down its explainability and transparency to preserve their intellectual property. In such cases, the court can protect those interests through its standard methodology for protecting confidential and proprietary information.[157]

b. Case-Based Reasoning

Rule-based systems require the input of experts to guide the development

151. *Id.*
152. *Id.* at 60.
153. *Id.* at 63.
154. *See id.* at 69 (illustrating the backward-chaining inference process in Figure 4.8).
155. *Id.* at 74.
156. *Id.*
157. *See generally* P. Kanagavel, *Intellectual Property Rights: A Comprehensive Overview*, 85 J. PAT. & TRADEMARK OFF. SOC'Y 663, 663–65 (2003) (discussing the different legal methods available for protecting intellectual property).

of the knowledge base.[158] Because of this it may be difficult to identify the correct rules (this is called the "knowledge-elicitation bottleneck").[159] This bottleneck arises due to several interacting factors: (1) the availability of the expert to work with the knowledge engineer; (2) the ability of the expert to articulate their knowledge to the knowledge engineer; (3) the ability of the knowledge expert to understand the problem and rules; and (4) the model of knowledge representation chosen by the knowledge engineer may not be able to represent the expert knowledge provided.[160] This bottleneck may make it practicably impossible to develop a set of expert rules.[161] In addition, in a rule-based system, we must know how to solve the problem posed. If we do not have that knowledge, we cannot use a rule-based system.[162]

In light of such difficulties, it may be useful to use a case-based methodology. In case-based reasoning, we start with a set of solved cases and infer from similarities between them and the queried case to arrive at a correct response even though we cannot articulate the rules for solving the problem.[163]

In a case-based system, the knowledge base is not a set of general if-then rules, but rather a set of specific cases representing specific instances of a problem and its solution.[164] That knowledge base is divided into a problem space, which provides the relevant information about the problem of the case, and the solution space, which provides the relevant solutions to those cases.[165]

The inference engine in a case-based system contains a number of subcomponents that work together. For example, the knowledge base must be structured so that relevant features of the case are identified.[166] These features can then be used to create an index which will aid in finding cases that are relevantly similar to the queried case.[167] Once an indexing methodology has

158. IAN WATSON, APPLYING CASE-BASED REASONING: TECHNIQUES FOR ENTERPRISE SYSTEMS 10 (1997).

159. *Id.*

160. *Id.*

161. *See* Andrew Young et al., *Parameterisation of Domain Knowledge for Rapid and Iterative Prototyping Knowledge-Based Systems*, 208 EXPERT SYS. WITH APPLICATIONS 1, 4 (2022) ("In the literature, it is agreed that the knowledge elicitation bottleneck is a major problem to overcome before being able to rapidly develop and deploy . . . expert systems.").

162. *See* Frederick Hayes-Roth, *Rule-Based Systems*, 28 COMM. ACM 921, 921–23 (1985) (providing an overview of rule-based systems and the different parts essential to their functioning).

163. *See* WATSON, *supra* note 158, at 46–48.

164. *See id.* at 13–14, 19.

165. *See id.* at 19.

166. *See id.* at 20.

167. *See id.* at 20–22.

been identified, the system must then adopt a retrieval methodology like nearest-neighbor retrieval.[168] In nearest-neighbor retrieval, the system plots the features of each case onto a graph to enable the system to note the distance between cases.[169] It then plots the queried case into this graph to determine which of our solved cases was closest.[170]

Relative to the knowledge base, case-based expert systems should easily meet the requirements of being explainable and transparent. We are all familiar with reasoning from a case (it is the primary method of teaching law students).

The added complexity of the inference engine demonstrates a fundamental problem with AI experts—the ability to understand them decreases as their sophistication increases.[171] This will not be as large of a problem for attorneys because they will have access to experts who will be able to understand the mathematical intricacies of the inference engine and identify any problems.[172] On the other hand, it creates substantial hurdles for the fact finder who must either understand intricate mathematics and complex logic or simply accept the explanation of the system given by experts without such an understanding.[173] We do not think this problem is insurmountable, but it is our first indication that the use of AI experts will likely require considerable oversight.

c. Bayesian Networks

Rule-based expert systems initially only engaged in deductive reasoning.[174] Unfortunately, deductive reasoning systems are not very capable when

168. *See id.* at 23.

169. *See id.* at 23–28.

170. *See id.*

171. *See generally* David Beer, *Why Humans Will Never Understand AI*, BBC (Apr. 7, 2023), https://www.bbc.com/future/article/20230405-why-ai-is-becoming-impossible-for-humans-to-understand ("There is a good chance that the greater the impact that artificial intelligence comes to have in our lives the less we will understand how or why.").

172. *See generally* Mason Ladd, *Expert Testimony*, 5 VAND. L. REV. 414, 417 (1954) (discussing how and when attorneys may access expert witnesses to help them and others within a courtroom understand complicated subject matter, especially in response to the "continuous and rapid progress of science").

173. *See generally id.* at 418–19 (describing how fact finders have—and will continue—to rely on expert witnesses for various legal issues "as a part of the growth of a scientific society in a complicated age").

174. ROBERT G. COWELL ET AL., PROBABILISTIC NETWORKS AND EXPERT SYSTEMS: EXACT COMPUTATIONAL METHODS FOR BAYESIAN NETWORKS 8 (Michael Jordan et al. eds., 2007) ("Originally, production systems involved only logical deductions.").

dealing with uncertainty.[175] In response, some rule-based systems created a "certainty factor" and assigned it to each if-then rule.[176] For example, we might have, "IF headache & fever THEN influenza (certainty 0.7)."[177] Such algorithms introduce probability theory into the expert system.[178] This addition to the knowledge base further complicates the inference engine, which must now incorporate algorithms for combining and interpreting such factors.[179] One such type of algorithm is based on Bayes's theorem, which is used to "mak[e] inferences in probabilistic expert systems."[180] Bayes's theorem is:[181]

$$P(A|B)=\frac{P(B|A) \times P(A)}{P(B)}$$

It works fairly simply but it is worthwhile to explain. Assume that *A* is the statement, "The aphids on my rose bush die." *P(A)* would represent how likely we think that the aphids on my rose bush die. If we have no other information, we would likely set that as a very unlikely thing with, perhaps, a probability of one in ten (0.1). *B* represents a new piece of information, in this case that "Lady bugs begin to eat the aphids." *P(B)* would represent the probability of lady bugs beginning to eat the aphids without any other information. We might set it as a slightly higher number than aphids dying—say one in five (or 0.2). Finally, we will need to know the value of *P(B|A)*—the probability of *B* occurring given that *A* occurred. With this new knowledge, we can ask what the probability that the aphids on my rose bush die given that lady bugs come to my rose bush to eat them (this is *P(A|B)*), or the probability of *A* occurring given *B* occurred. *P(A)* is called the prior probability (which represents our belief about the probability of *A* occurring before observing the additional evidence) and *P(A|B)* is posterior probability (which represents our belief about the probability of *A* occurring with the additional evidence from *B*).[182]

175. *Id.* at 8–9.
176. *Id.* at 9.
177. *Id.*
178. *See id.* at 10.
179. *Id.*
180. *Id.* at 14–21.
181. *Id.* at 14.
182. *Id.*

If this short, high-level discussion has caused your eyes to glaze over, think about the effect it will have on the average fact finder. The vast majority of people have little intuitive understanding of probability theory and will have difficulty understanding this feature of the Bayesian system.[183]

d. Fuzzy Logic

As we have seen, Bayesian Networks use probability theory and statistics to manage uncertainty.[184] Fuzzy expert systems also address uncertainty but do so from a non-statistical framework focusing on fuzzy sets (the fuzzy knowledge base) and fuzzy logic (the fuzzy inference engine).[185]

A fuzzy set is composed of elements that are assigned a grade of membership in the class the set is intended to represent.[186] For example, assume we have a fuzzy class, <High GPA>, intended to capture the idea of a student having a high GPA. That class will have a series of elements representing actual GPAs (e.g., the elements {2.9, 3.2, 3.8}). In a fuzzy set, we then associate a grade of membership (how well that element fits into <High GPA>) with each GPA, where a higher membership grade means the associated GPA is more likely a member of the set. For example:

GPA	Membership grade
2.9	0.1
3.2	0.4
3.8	0.8

In a fuzzy expert system, the knowledge base is composed of a collection of fuzzy rules, membership functions, and associated information that represents the (imprecise) expertise and domain knowledge of the system.[187] Fuzzy rules capture the expert knowledge and define the relationships between

183. *See* Amos Tversky & Daniel Kahneman, *Judgement Under Uncertainty: Heuristics and Biases*, 185 SCIENCE 1124, 1124–30 (describing the sources of difficulty in understanding probability: representativeness, availability, and adjustment and anchoring).

184. *See* COWELL ET AL., *supra* note 174, at 14 (discussing how Bayes's theorem is the "basic tool for making inferences in probabilistic expert systems," and how it permits one to account for random variables).

185. *See* Abraham, *supra* note 136, at 912.

186. *See id.*

187. *See id.* at 912–13.

inputs and outputs in the system.[188] Usually a fuzzy rule consists of an antecedent (input conditions) and a consequent (output action) of the form: "If *A* is *low* and *B* is *high* then *X* [is] medium."[189] Membership functions are used to describe the degree of membership or membership grades of values in fuzzy sets.[190] They define the fuzzy boundaries and shape of the linguistic variables used in the fuzzy rules.[191] Each input and output variable in the fuzzy expert system is associated with one or more membership functions, which assign membership grades to values based on their degree of similarity to the fuzzy sets.[192] The fuzzy inference engine then applies fuzzy logic to the fuzzy knowledge base to draw conclusions based on the input data.[193]

Because the input into a fuzzy expert system may be crisp (meaning having a truth value of 1) or fuzzy, and it may also be desirable for the output (which normally will be fuzzy) to be crisp, a fuzzy expert system has two additional features (compared to standard expert systems): a fuzzification interface and a defuzzification interface.[194] This gives us a slightly different graphical representation than other expert systems:

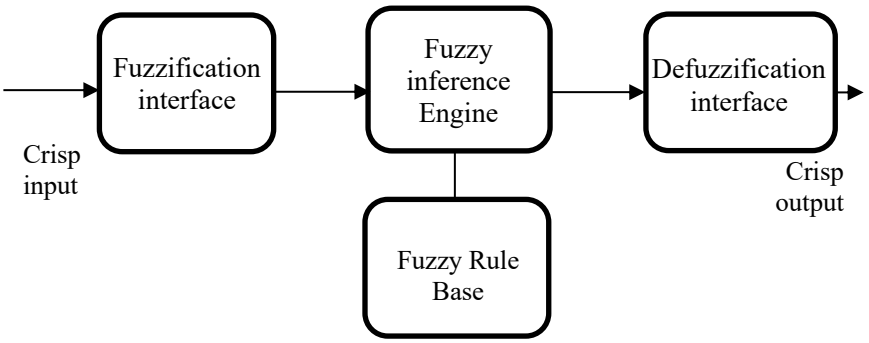

Figure B. Architecture of a fuzzy expert system.[195]

188. *See id.* at 918.
189. *Id.* at 912.
190. *See id.*
191. *See id.* at 912–13.
192. *See id.*
193. *See* CONSTANTIN VIRGIL NEGOITA, EXPERT SYSTEMS AND FUZZY SYSTEMS 95–112 (Alan Apt & Antonio Padial eds., 1985) (providing a general discussion of rules of inference for fuzzy systems).
194. *See* Abraham, *supra* note 136, at 912–13 (discussing the fuzzification and the defuzzification interfaces).
195. *See id.* at 913.

The fuzzification interface fuzzifies the input variables "whereby the membership functions defined on the input variables are applied to their actual values[] to determine the degree of truth for each rule antecedent."[196] Because the output from the fuzzy interference engine is a fuzzy set, the defuzzification interface "extracts the crisp output that best represents the fuzzy set."[197]

e. Transparency and Explainability in Expert Systems

As we have seen, as expert systems become more sophisticated, they become more difficult to understand. Because attorneys will have access to experts who do understand the underlying mechanics, this higher sophistication will not impede them beyond the financial costs of using such an expert.[198] But there is a more pressing problem for the fact finder who is unlikely to be able to understand the sophisticated mathematics and logic (fuzzy set theory, probability and statistics, deductive logic) required to understand the system (a requirement of both explainability and transparency). As such, it will be more difficult, relative to the fact finder, to meet the other explainability requirements of fidelity and sufficiency. Despite these concerns, we should recognize that expert systems are transparent and explainable, and therefore are, in principle, open-box systems.[199]

2. Machine Learning Systems

In expert systems, we develop a knowledge base and inference engine to draw conclusions based on a query from the user.[200] This requires us to already know something about the regularities (rules) we are attempting to

196. *Id.* at 912.

197. *See id.* at 913.

198. *See* Cindy Siegel, *Will AI Technology Generate the Next Big Wave of Litigation?*, WIT LEGAL (Aug. 11, 2023), https://www.witlegal.com/insights/article/ai-technology-experts-for-litigation/ (discussing how lawyers can use AI experts); *see also* Zachary Crockett, *The Lucrative Economics of Expert Witnesses*, THE HUSTLE (Feb. 9, 2024), https://thehustle.co/the-lucrative-economics-of-expert-witnesses/ (outlining expenses associated with expert witnesses).

199. Simon J. Preis, *Are Expert Systems Dead?*, MEDIUM (Mar. 16, 2023), https://towardsdatascience.com/are-expert-systems-dead-87c8d6c26474 ("The result of an ES is always transparent."); Abraham, *supra* note 136, at 918 (describing expert systems as expressing knowledge in "easy-to-understand" rules).

200. Martin Stytz, *The Tale of Two AIs: GPT v. Expert Systems*, ITC FED., https://itcfederal.com/news/the-tale-of-two-ais-gpt-vs-expert-systems/ (last visited Feb. 6, 2024) (explaining knowledge bases and inference engines in expert systems).

model and how we should derive conditions for them. That means that expert systems are not useful where we don't understand the underlying regularities enough to write them down. This is a problem because for many important expert systems like vision, speech, translation, and robotics, we have not been "able to devise very good algorithms despite decades of research beginning in the 1950s."[201]

On the other hand, if there is sufficient data, we can use machine learning to model the regularities in the data even if we don't know what they are.[202] At a high level, machine learning can be visualized as follows:

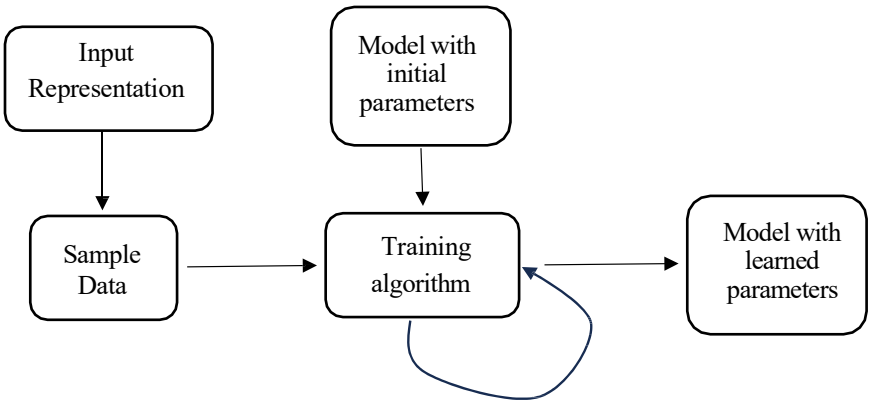

Figure C. Machine Learning Process

Where the input representation is the attributes of the data that we believe are relevant to the problem constitutes the input representation.[203] For example, if we are seeking to develop a system to estimate the value of a used car, the attributes might include the vehicle's age, make, and mileage.[204] The sample data is the subset of the data we are using to train the system.[205] We use statistical sampling so that once we complete the initial training, we can use

201. ETHEM ALPAYDIN, MACHINE LEARNING x (2016).
202. *See* ALPAYDIN, *supra* note 96, at 1–2.
203. *See id.* at 21 ("Note that when we decide on this particular *input representation*, we are ignoring various other attributes as irrelevant.").
204. *See id.* at 36.
205. *See id.* at 24 ("What we have is the training set X, which is a small subset of all possible x.").

the unsampled data to validate the model. The model, in general, is the structure or template we believe represents the relationship between the input and the output.[206] For example, if we believe the price of a used car is a linear function of the mileage, we can represent that data as with the model $Price = \alpha Mileage + \beta$, where α is the effect of mileage on price and β is a constant that represents the general floor for price.[207] What we want the algorithm to uncover for us is the actual value of the parameters of the model (the actual value of α and β).[208] The training algorithm begins with the model plus whatever we set the parameters to be initially.[209] It then uses the data to learn the value of α and β as represented in that data.[210] Once it has completed that process, we have a model with fixed parameters that we can confirm as accurate by using the unsampled data to see if the model correctly predicts the price of a used car in the unsampled data.

a. Models

In machine learning, we start with a hypothesis about the structure of how the input data is related to the output data.[211] When we do not know the true relationship between the input and output, there are many hypotheses that could fit the data.[212] When we choose a particular hypothesis and associated model, we create inductive bias—a set of assumptions about the phenomena under study.[213] Our goal in model training is to use the data to adjust the parameters of the model so that it correctly predicts the right output for the input (called "generalization").[214] There are two types of errors we can make in creating the model. Our hypothesis and model may be insufficiently complex and therefore underfit the data, or the hypothesis and model may be too

206. *See id.* at 8 (discussing that a model helps explains "the process underlying the data").

207. *See id.* at 10 (explaining a similar equation).

208. *Id.* (describing how the model optimizes parameters to get the closest estimates).

209. *Id.* at 3 ("We have a model defined up to some parameters, and learning is the execution of a computer program to optimize the parameters of the model . . . .").

210. *Id.* at 10.

211. *Id.* at 23 (describing how to create a hypothesis class).

212. *Id.* at 23–24 ("Though the expert defines this hypothesis class, the value of the parameters are not known, that is, though we choose [a hypothesis], we do not know which particular [hypothesis] is equal, or closest, to [correct].").

213. *Id.* at 38.

214. *Id.* at 38–39 ("How well a model trained on the training set predicts the right output for new instances is called *generalization*.").

complex and therefore overfit the data.[215] Both underfitting and overfitting result in a set of parameters that is sub-optimal for generating the correct output.[216]

Earlier, we modeled estimating the price of a used car using a linear equation.[217] In that case, we had a single variable—mileage. But that might underfit the data, and we might prefer to have a model with multiple variables. Perhaps it would be better if we added in the variable of brand. In addition, we assumed that there was a linear relationship between mileage and price—meaning that for every fixed increase in mileage (e.g., 5,000 miles) there is a fixed decrease in price (e.g., $100). But this relationship might not be linear. For example, the amount the price goes down might be much greater the higher the mileage.

Moreover, if mileage and brand are not linearly related to the price of the used car (a fixed decrease in mileage equals a fixed decrease in price), we might need to use a nonlinear function. For example, it could turn out that as mileage on a used car increases its price decreases at a faster rate. We can see such a result in the following diagram:

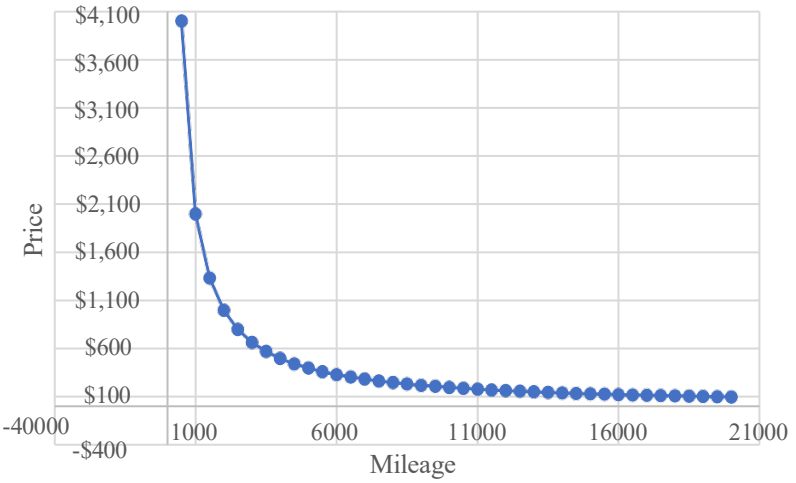

Diagram A: Nonlinear Relationship Between Mileage and Price

215. *Id.* at 39.
216. *Id.* (outlining how underfitting and overfitting lead to poor generalization).
217. *See supra* text accompanying notes 203–210 (using a linear equation to estimate the price of a used car).

To provide a broader understanding of how we might model a system, we provide two further examples of modeling. First, we look at a system designed to enable a person to find the shortest route between two points. Second, we explain neural networks.

Let us assume we have a warehouse with six destinations: shelf 1, shelf 2, assembly table, delivery dock, front office, and break room.[218] We represent each of these different destinations with a different letter: *A*, *B*, *C*, *D*, *E*, and *F*. We want to develop a system that will show the shortest route between any two points. Let us assume these points are connected in the following manner, with the circles representing the various destinations and the lines representing the path from a destination to another:

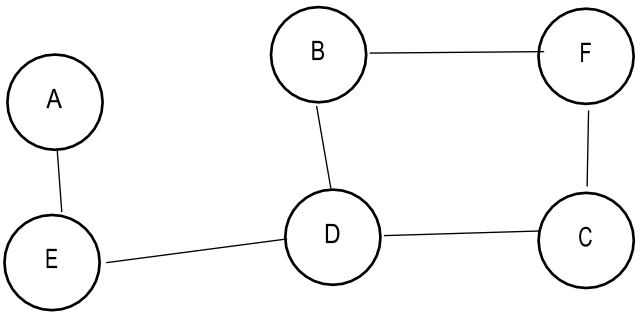

Diagram B: Warehouse Destination Connections

This problem cannot be easily (or intuitively) represented by an equation. Instead, we represent it by a matrix:

218. DENIS ROTHMAN, ARTIFICIAL INTELLIGENCE BY EXAMPLE 7–25 (Tushar Gupta et al. eds., 2d ed. 2020) (providing the model from which this one was derived).

	A	**B**	**C**	**D**	**E**	**F**
A	0	0	0	0	α	0
B	0	0	0	α	0	α
C	0	0	0	α	0	α
D	0	α	α	0	α	0
E	α	0	0	α	0	0
F	0	α	α	0	0	0

Diagram C: Matrix Model of Warehouse Destination

In this matrix, the first column of each row represents one of the six possible current locations. The top row of letters represents a destination. In our model, we put a zero if you cannot move from the current location to the destination and some number α which falls into the range (0,1).[219] We have also assumed that because this problem is about going from one destination to another, we will use a 0 for going from one destination directly to the same destination.[220] For example, let us assume we are in destination *D*. When we look at the associated row, we see there is a non-zero value in the columns associated with destinations *B*, *C*, and *E*. That means you can go to those destinations from destination *D*. Importantly it is the matrix structure that is the model. The values associated with each cell (row/column pair) are the parameters which, in this case, we have initially set to 0 and α. And we change these through training on data sets.

Another model that is often used in AI is the neural network.[221] The neural network model is based on our understanding of the functioning of the brain.[222] The brain is considered to have two low-level features: the neuron (which is the processing unit) and the synapses which connect the neurons

219. *See id.* at 11, 18 (providing a similar matrix). In this example, we have used a single value for α, but we could also set each row and column to a different value (having multiple α's). *See id.* at 18 (demonstrating a matrix with multiple α's).

220. *Id.* at 11 (explaining how zero is used for locations); *see also supra* Diagram C and accompanying text (describing how the problem is about going from one destination to another).

221. ALPAYDIN, *supra* note 96, at 267–68 (introducing neural networks).

222. *Id.* at 267 ("Artificial neural network models . . . take their inspiration from the brain.").

(the parameters).[223] Unlike traditional computer systems where instructions are executed serially (one by one), neurons act in parallel—each neuron can process its information at the same time as any other neuron.[224] It is believed that neurons engage in the processing and that memory (and learning) arises in the synapses.[225]

One way of modeling a neuron is through a perceptron:

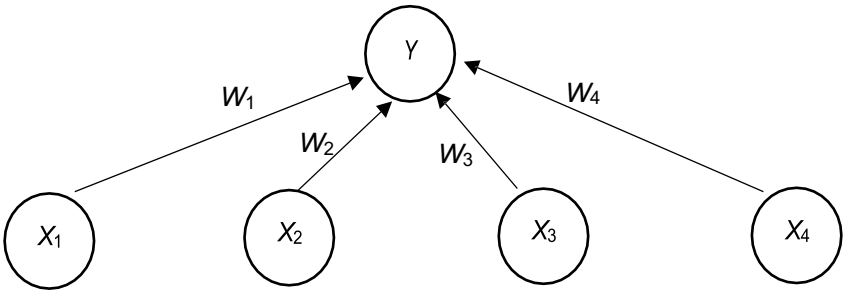

Diagram D: Single Neuron Perceptron

Where Y is the output unit, each X_i is an input unit, and each W_i is the weight of the connection from the related input to the output.[226] Just as in our matrix model, the values of the weights (the parameters) are initially set to some value α (or some set of values α_1 to α_i) and then modified through the learning process.[227] In using the perceptron, the user provides it with the relevant input for each input unit.[228] The output unit then takes a weighted sum of the weights and inputs to produce some output.[229] For example, if we set the input units and weights as follows, the value of Y will be 0.5:

223. *Id.* at 267–68.
224. *Id.* at 267–68, 270 (discussing how neurons work in parallel and parallel processing).
225. *Id.* at 267–68.
226. *Id.* at 271.
227. *See id.* at 272 (explaining that the weights are the parameters of the systems and the learning process generates the correct outputs).
228. *Id.* at 271 (describing perceptrons and their inputs).
229. *Id.* ("[T]he output, *y*, in the simplest case is a weighted sum of the inputs . . . .").

		Weighted sum
$X1 = 1$	$W1 = 0.3$	0.3
$X2 = 0$	$W2 = 0.5$	0
$X3 = 1$	$W3 = 0.2$	0.2
$X4 = 0$	$W4 = 0.9$	0
		$Y = 0.5$

A perceptron may be combined into multiple parallel perceptrons:

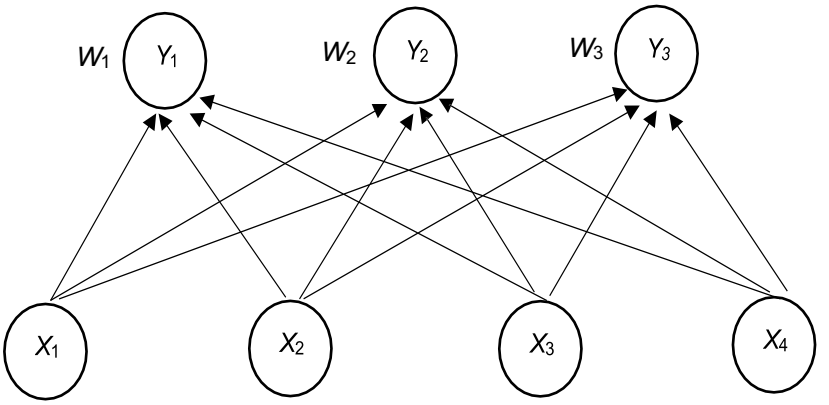

Diagram E: Multiple Parallel Perceptrons

Here we still have four input units (X_1 . . . X_4) but now each input unit connects to each of three output units (Y_1, Y_2, Y_3). We represent the sum of the input weights for each output unit by the weighted average W_i.[230] For example, W_1 represents the weighted sum of the inputs from X_1, X_2, X_3, and X_4.[231]

In the models we have created thus far, there is a direct connection

230. *Id.* at 271–73 (explaining the weighted average from the input and output weights),
231. *Id.*

between the input and the output.[232] This implies that there is a linear relationship between the input and the output—a one-to-one correspondence.[233] But sometimes that is not the case.[234] To represent more complex and nonlinear relationships our perceptron model will have a hidden layer:[235]

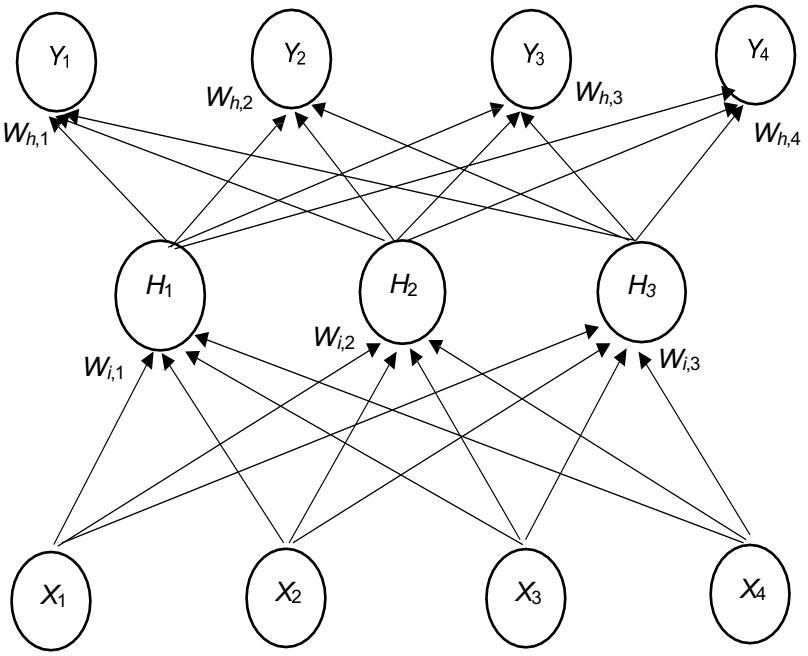

Diagram F: Multiple Parallel Perceptrons with a Single Hidden Layer

Where X_i is input unit i, H_i is hidden unit i, Y_i is output unit i, $W_{i,j}$ is the weighted sum of the input units' connections to the hidden unit H_j, and $W_{h,k}$ is the weighted sum of the hidden units' connections to Y_j.[236] This neural

232. *See supra* text accompanying notes 203–231 (describing previous models).

233. ALPAYDIN, *supra* note 96, at 272 ("Thus this perceptron with one input and one output can be used to implement a linear fit.").

234. *Id.* at 279 ("[*M*]*ultilayer perceptrons* (MLP) can implement nonlinear discriminants and, if used for regression, can approximate nonlinear functions of the input.").

235. *Id.* (explaining that networks with hidden layers can be used for nonlinear regression).

236. *Id.* at 279–81 (explaining inputs and outputs in hidden layers).

network represents an encoder where the outputs equal the inputs and the number of hidden units is less than the number of outputs/inputs.[237] Such an encoder can be used to find a simpler structure to represent the input/output.[238]

If we believe that our input has multiple features relevant to the output, we may want to use a neural network with multiple hidden layers. A network with many hidden layers is called a deep neural network where "successive hidden layers correspond to more abstract representations until we get to the [final] output layer."[239] For example, a deep learning network attempting to identify objects in pictorial form might have a hidden layer for edges, another layer might be for corners, and yet another might be to identify arcs.[240] These levels feed into each other until they end in the output—what the picture represents.[241]

b. Types of Learning Algorithms

Thus far we have only discussed how we can structure a model to represent data. Next, we must understand how the AI system uses that model and the data to learn things. Machine learning can be defined as "programing computers to optimize a performance criterion using example data or past experience."[242] To do so we must postulate a model with an initial set of parameters and then allow the learning algorithm to review the data and optimize those parameters.[243] We will look at three basic categories of machine learning systems: supervised learning, unsupervised learning, and reinforcement learning.

In supervised learning, we have the input and the output we desire, and we want the learning algorithm to identify the relationship between the two.[244] At its core, a supervised learning algorithm has four components: (1) a sample of the input data with associated "correct" output, (2) a model with associated parameters, (3) a loss function that computes the difference between our current estimate of the output and the expected output, and (4) an optimization

237. *Id.* at 303.
238. *Id.*
239. *Id.* at 307–08.
240. *Id.* at 308 (explaining the hidden layers that might be present).
241. *Id.* at 308–09 (describing how the abstract layers result in the correct output).
242. *Id.* at 3.
243. *Id.*
244. *Id.* at 11, 41 (describing supervised learning).

procedure that minimizes the error between the predicted output and the expected output.[245] Loss functions include linear regression,[246] logistic regression,[247] k-nearest neighbors,[248] decision tree,[249] support vector machines,[250] random forests,[251] and naïve Bayes.[252]

In unsupervised learning, we provide a model with the input data but without the associated output data.[253] We then use an unsupervised learning algorithm to discover regularities in the data.[254] A primary unsupervised learning technique is called "clustering" where the data attempts to identify groupings within the data.[255] Clustering algorithms come in a variety of forms including k-means clustering, expectation-maximization, spectral clustering, and hierarchical clustering.[256] Other methods of unsupervised learning include anomaly detection,[257] association rule learning,[258] and generative adversarial networks.[259]

Reinforcement learning has been "called 'learning with a critic.'"[260] In it, the system acts or draws a conclusion.[261] When it does so well, the components of the system involved in that choice are strengthened, and if it fails, the components associated with that failure are weakened.[262] Reinforcement learning algorithms include Q-learning,[263] Deep Q-Networks,[264] Policy

245. *Id.* at 41–42.
246. *Id.* at 79.
247. *See id.* at 250–57.
248. *See id.* at 190–92.
249. *See id.* at 213.
250. *See id.* at 353–54.
251. *See id.* at 234–35.
252. *See id.* at 397–99.
253. *Id.* at 11.
254. *Id.*
255. *Id.*
256. *Id.* at 165–82.
257. *Id.* at 207.
258. *Id.* at 56–59.
259. *Id.* at 396–99.
260. *Id.* at 518.
261. *Id.* at 517.
262. *Id.* at 518.
263. Sergio Spanò et al., *An Efficient Hardware Implementation of Reinforcement Learning: The Q-Learning Algorithm*, 7 IEEE ACCESS 186340, 186341 (2019).
264. Hado van Hasselt et al., *Deep Reinforcement Learning with Double Q-Learning*, 30 PROC. THIRTIETH AAAI CONF. ON ARTIFICIAL INTELLIGENCE 2094, 2095 (2016).

Gradient methods,[265] Actor-Critic methods,[266] and Monte Carlo Tree Search.[267]

3. A General Model of an Artificial Intelligence System

Based on the discussion above, we can provide a common structure for an AI system based on a series of models.

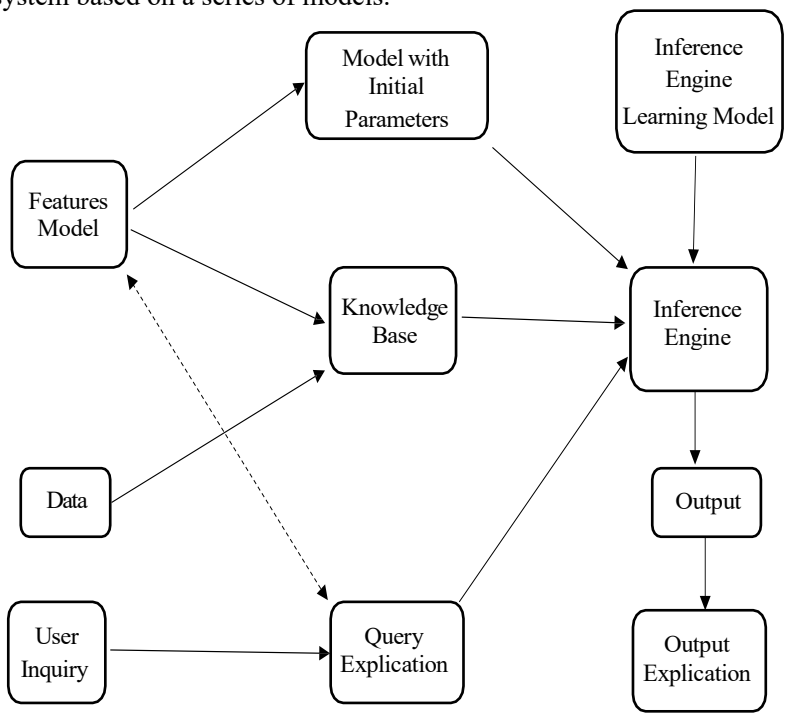

Diagram G: Generic Depiction of an AI System

265. Richard S. Sutton et al., *Policy Gradient Methods for Reinforcement Learning with Function Approximation*, 12 ADVANCES NEURAL PROCESSING SYS. 1057, 1057–60 (2000).

266. Andrea Zanette et al., *Provable Benefits of Actor-Critic Methods for Offline Reinforcement Learning*, 35 CONF. ON NEURAL INFO. PROCESSING SYS. 1, 2 (2021).

267. Tom Vodopivec et al., *On Monte Carlo Tree Search and Reinforcement Learning*, 60 J. ARTIFICIAL INTELLIGENCE RES. 881, 881–82 (2017).

Before we can build either a model of the system or a knowledge base, we must know what features in the data are relevant. The features model provides us with that information.[268] For example, if we are looking at a visual identification system we might be focusing on the color and intensity of a pixel in some array.[269] If we are using a case-based expert system, we will need to identify the features of the case that are potentially relevant to the inquiry.[270]

Once we have a features model, we can create the knowledge base which will contain the original data but represented by the relevant features of that data.[271] In our case-based expert system, this would be the set of cases represented in terms of the relevant features of the cases.[272] In a visual identification system, this would be the actual array of pixels and associated features (intensity and color).[273] Similarly, using the features model, we can identify the model we want the inference engine to use.[274] This might be a fuzzy expert system, an equation, an array, or a neural network.[275] As part of this process, where applicable, we would choose the initial values of all parameters.[276]

We also need a model of how the inference engine will work. For example, in a rule-based expert system we choose between forward chaining or

268. *See* Matthew McMullen, *What Are Features in Machine Learning and Why Is It Important?*, MEDIUM (July 15, 2019), https://cogitotech.medium.com/what-are-features-in-machine-learning-and-why-it-is-important-e72f9905b54d (explaining a features model in machine learning); *see also* Kevin Stumpf, *What Is a Feature Platform for Machine Learning*, TECTON (Oct. 17, 2023), https://www.tecton.ai/blog/what-is-a-feature-platform/ (describing a features model).

269. *See* Gaudenz Boesch, *Image Recognition: The Basics and Use Cases (2024 Guide)*, VISO.AI, https://viso.ai/computer-vision/image-recognition/ (last visited Feb. 8, 2024) (discussing visual identification models in AI).

270. *See Case Based Reasoning*, LARK (Dec. 23, 2023), https://www.larksuite.com/en_us/topics/ai-glossary/case-based-reasoning (explaining case-based reasoning and what is relevant).

271. *See* Abraham, *supra* note 136, at 910 (discussing knowledge bases as "stor[ing] all relevant information, data, rules, cases, and relationships used by the expert system").

272. *See Case Based Reasoning*, *supra* note 270 (explaining how case-based systems use relevant cases).

273. *See* Boesch, *supra* note 269 (discussing AI and visual identification systems).

274. *See Inference Engine*, AUTOBLOCKS, https://www.autoblocks.ai/glossary/inference-engine (last visited Feb. 9, 2024) (explaining inference engines).

275. *See id.* (describing examples of inference engines); *see also* Abraham, *supra* note 136, at 912 (discussing inference engines in fuzzy expert systems).

276. *See Inference Engine*, *supra* note 274 (explaining that inference engines use relevant facts set by the knowledge base to create outputs); *see also* ALPAYDIN, *supra* note 96, at 3 (outlining how machine learning uses models defined by set parameters).

backward chaining.[277] In a machine learning system, we would determine whether we are using supervised, unsupervised, or reinforcement methodology.[278] Within that methodology we would then identify what particular algorithm we would use.[279] For example, if we choose reinforcement learning, then we would use Q-Learning, Deep Q-Networks, Policy Gradient methods, Actor-Critic methods, or Monte Carlo Tree Search.[280] Once we understand how the inference engine will work, we will need to create an inference engine that will implement that approach given our knowledge base.[281]

In order to get a response from the system, the user must make a query.[282] The user's query will likely not be in a format compatible with the structure of the knowledge base and inference engine, so it will have to be processed to make it compatible.[283] For example, if we ask a fuzzy expert system, the diagnosis of a patient with a fever of 102.5°, chills, and a sore throat, we would need to take the very crisp 102.5° and convert it to a fuzzy concept (e.g., high fever, medium fever, low fever). Similarly, if we ask a machine learning system if there is a dog in this picture, we may need to take a physical picture and digitalize. The query explication system is how we do that.[284]

Finally, the system will want to provide us with an answer to our query. We use an output explication system to convert, where necessary, the AI

277. Abraham, *supra* note 136, at 910 (explaining forward chaining and backward chaining as commonly used in rule-based expert systems).

278. ALPAYDIN, *supra* note 96, at 11–14 (discussing supervised learning, unsupervised learning, and reinforcement learning as examples of machine learning systems).

279. *See, e.g.*, Robert Moni, *Reinforcement Learning Algorithms—An Intuitive Overview*, MEDIUM (Feb. 18, 2019), https://smartlabai.medium.com/reinforcement-learning-algorithms-an-intuitive-overview-904e2dff5bbc (explaining algorithms used in reinforcement learning).

280. *Id.* (outlining reinforcement learning algorithms and methods).

281. Abraham, *supra* note 136, at 910 ("The purpose of the inference engine is to seek information and relationships from the knowledge base . . . .").

282. *See Query Language in AI*, LARK (Dec. 22, 2023), https://www.larksuite.com/en_us/topics/ai-glossary/query-language-in-ai (explaining how a query is "the process by which machines interpret and respond to human language inputs"); *see also* Yufeng G, *The 7 Steps to Machine Learning*, MEDIUM (Aug. 31, 2017), https://towardsdatascience.com/the-7-steps-of-machine-learning-2877d7e5548e (outlining the steps to machine learning).

283. *See* Yufeng G, *supra* note 282 (explaining data preparation); *see also* Abraham, *supra* note 136, at 912 (describing expert system development and how knowledge often needs to be entered in a specific format).

284. *See Query Language in AI*, *supra* note 282 (explaining how queries are how machines respond to humans); *see also Query*, APP LOVIN, https://www.applovin.com/glossary/query/#:~:text=A%20query%20is%20a%20special,u (last visited Feb. 9, 2024) (describing a query as "a question written in a very specific format so that computers can understand and respond to it").

system's information into a format understandable by the user.[285]

4. Meeting the Requirements of Explainability and Transparency

Now that we have a basic understanding of AI systems and a generic model of such a system, we can see the difficulties with using such a system's output as evidence. An AI system has ten distinct parts: features model, model with initial parameters, inference engine learning model, inference engine, knowledge base, data, user inquiries, query explication, output, and output explication.[286] Each of these distinct parts must be both transparent and explainable.

This will not be as simple as it might appear. One example is data. The data used in a system is not simply predefined as data relevant to that system. The developer must make a decision as to what data they will use and what data they will not. For example, if we want to build a machine learning facial identification system, we will need to have a lot of pictures of faces to train the system.[287] To get those pictures we will need to identify a source of pictures.[288] But what source? We could trawl Facebook, Instagram, or LinkedIn to get our set of pictures. Given that the demographics of each of those systems are different, the source we choose could affect how well our facial identification system works for certain categories of faces.[289] If we use a system that is predominantly male and White, our system might be great at identifying White males and terrible at identifying everyone else.[290] Even if our population of photos is not biased as to race and gender, the way we take a sample from those photos might introduce bias. As such, the litigator must examine not merely the data itself, but the underlying decisions relating to the collection and creation of the data.

285. *See Query Language in AI*, *supra* note 282 (explaining how queries help give answers tailored to users); *see also* Yufeng G, *supra* note 282 (outlining the steps to machine learning).

286. *See supra* Section III.D.3 (outlining and explaining the ten parts of an AI model).

287. *See Machine Learning and Face Recognition*, PXL VISION (June 7, 2022), https://www.pxl-vision.com/en/blog/machine-learning-and-how-it-applies-to-facial-recognition-technology (explaining that facial recognition works by comparing faces to a database, so numerous pictures are needed).

288. *See id.* (noting that facial recognition works off pictures in a database).

289. Sidney Perkowitz, *The Bias in the Machine: Facial Recognition Technology and Racial Disparities*, MIT SCHWARZMAN C. COMPUTING (Feb. 5, 2021), https://mit-serc.pubpub.org/pub/bias-in-machine/release/1 (explaining how the demographics of databases can impact how well facial recognition works).

290. *Id.* (describing how a lack of nonwhite faces can impact the facial recognition software's performance).

IV. APPLYING CURRENT RULES TO FORENSIC AI EVIDENCE

The challenges of AI development undoubtedly will require new ethical, legal, and regulatory change over time, and all of those changes will be necessary to appropriately constrain AI usage in law enforcement and as evidence.[291] Certain scholars, such as Erin Murphy advocating for structural quality control mechanisms,[292] Andrea Roth evaluating all types of machine-based testimony,[293] or Edward Cheng and G. Alexander Nunn advocating for enhanced discovery in cases involving AI evidence,[294] have begun the conversation on appropriate legal responses to technological development. But as those conversations continue, judges in the courtroom must apply the current rules to the technological evidence offered in court today. We therefore offer a roadmap for current judicial assessment of the management of AI evidence in court, under both the rules and constitutional limitations. The application of the rules to AI largely depends on the computational methodology underlying the conclusions reached. As discussed in detail below, only certain types of AI systems will meet the current standards for reliability and confrontation, while others will fail. These applications will, by necessity, lead to our conclusions about future policy development as well.

Before we get to the application, however, a quick reminder of the impediments to the admission of AI-generated evidence is in order. If evidence is being offered based on "scientific, technical, or other specialized knowledge," then under Federal Rule of Evidence 702, that evidence must be screened by the judge to ensure proper reliability.[295] In federal court, such reliability screening necessitates an assessment of the underlying methodology, because as *Daubert* itself made clear, the "focus, of course, must be solely on principles and methodology, not on the conclusions that they generate."[296] In assessing the methodology, courts can examine various considerations, among them whether the method has been tested, subject to peer review, has a known rate of error or standards, and general acceptance among the

291. *See* Roth, *supra* note 18, at 1292 (discussing issues facing the use of AI in the criminal justice system).

292. Erin Murphy, *The Mismatch Between Twenty-First-Century Forensic Evidence and Our Antiquated Criminal Justice System,* 87 S. CAL. L. REV. 633, 657–60 (2014).

293. Roth, *supra* note 5, at 1972; *see also* Roth, *supra* note 18, at 1245.

294. Edward K. Cheng & G. Alexander Nunn, *Beyond the Witness: Bringing a Process Perspective to Modern Evidence Law*, 97 TEX. L. REV. 1077, 1079–82 (2019).

295. FED. R. EVID. 702.

296. Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 595 (1993).

relevant scientific community.[297] A judge addressing reliability, however, can assess any factor or issue that allows an overall determination of whether the method is scientifically valid and reliable, such that the proponent has not left the judge with "too great an analytical gap between the data and the opinion proffered."[298]

When faced with evidence based on AI technology, then, a court must assess the computational methodology using these prefabricated tools which, based on their creation in reaction to widely divergent circumstances, may not be a great fit to the task at hand.[299] Yet, the underlying need to assess reliability of the method is at the core of ensuring legitimacy of the decision-making process in court, and thus legitimacy of its outcomes.[300] Furthermore, the right of a criminal defendant to confrontation, under the Sixth Amendment, will provide additional safeguards to the admission of unreliable evidence, although in this case it may not significantly vary from the initial reliability screening.[301]

Applying the reliability standard to the computational methods discussed in Part III, we conclude that expert systems are likely to be admissible as reliable, although we suggest particular caution for judges facing expert systems that utilize probability, statistics, fuzzy logic, and other sophisticated mathematics in the system. Under the same set of rules, machine learning and neural network AI should not be admitted because its methodology will fail the reliability gatekeeping, as even a system that creates valid outcomes will always have "too great an analytical gap" as described by *Joiner*.[302]

A. Application to Rule and Case-Based Expert Systems

In Part III, we explored the architecture of different AI systems based on their computational methodology, beginning with rule- and case-based expert

297. *Id.* at 593–94.

298. Gen. Elec. Co. v. Joiner, 522 U.S. 136, 146 (1997).

299. *See Daubert*, 509 U.S. at 592–95 (discussing ways courts assess methodology, including peer review and known rate of error, to ensure proper credibility).

300. *See, e.g.*, Jane C. Moriarty, *The Inscrutability Problem*: *From First-Generation Forensic Science to Neuroimaging Evidence*, 60 DUQ. L. REV. 227, 230–37 (2022) (comparing the issue of AI to traditional forensic sciences and advocating for a reliability-based assessment to admit AI evidence).

301. *See infra* Section IV.A (discussing in detail the applicability of the Confrontation Clause to machine learning and the current state of debate, and concluding that it does apply); *see also infra* Section V.B (suggesting additional work in the area is necessary and additional cases to decide in future).

302. *Joiner*, 522 U.S. at 146.

systems. Rule-based systems contain a knowledge base of facts relevant to the subject at hand, and then heuristics or rules used to solve problems in a particular domain, often in the form of "if . . . then . . . ." expressions.[303] Case-based systems, on the other hand, start with a set of solved cases and infer from similarities between them and the queried case to reach a correct response to the new inquiry.[304] What both methodologies have in common, as reviewed in Part III, is that they are, at a fundamental level, both explainable and transparent, in that they provide a user with a qualitative understanding of the connection between the input and the output,[305] while also being open and accessible for review.[306]

The key to assessing reliability in the case of AI evidence is to assess the ability to review the computation that leads to a specific conclusion.[307] Thus, when presented with evidence created using AI expert systems, a court should require expert testimony to trace the methodology that the system used to reach a specific conclusion.[308] A judge should be able to connect the dots from the general system architecture to its application in a specific case and be convinced that the methodology represents a valid scientific process.[309] If so, and due to the architecture of the systems in question, the evidence will likely satisfy the gatekeeping standard of *Daubert* and be admitted.[310] On the other hand, if the proponent is unable to connect the dots, explain the system methodology, or discuss how a specific conclusion is reached, whether due to lack of expert testimony, vague expert testimony, or leaving "too great an analytical gap" for the judge, then gatekeeping has not been satisfied and the evidence should be excluded.[311]

Of course, in the subset of cases in which AI evidence is being offered against a criminal defendant, the Confrontation Clause could affect

303. *In re* Lockwood, 679 F. App'x. 1021, 1028 (Fed. Cir. 2017); Synopsys, Inc. v. Ricoh Co. Ltd., Nos. C 03–2289 MJJ, C 03–4669, 2005 WL 6217119, at *11 (N.D. Cal. Apr. 7, 2005); *see supra* text accompanying notes 138–142.

304. WATSON, *supra* note 158, at 46–48; *see also supra* text accompanying notes 158–170.

305. *See* Ribeiro et al., *supra* note 121, at 1135; *see also supra* text accompanying notes 132–135.

306. *See* Weller, *supra* note 125, at 23–27.

307. Roth, *supra* note 5, at 2047–48.

308. *See id.* at 1981–82 (discussing how experts should testify to methodology).

309. Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 592 (1993). Convinced, at least, to the required preponderance of the evidence standard. *Id.*

310. *Id.* at 592–95 (outlining the factors for assessing expert testimony).

311. *Id.* (discussing the standard necessary for admitting expert evidence); Gen. Elec. Co. v. Joiner, 522 U.S. 136, 146 (1997) (explaining the court can determine there too great a gap between the data and the opinion offered).

admissibility as well. Commentators and scholars have debated the applicability of the Confrontation Clause to machine-based evidence, including AI, in recent years.[312] Andrea Roth, in her article *Machine Testimony*, reviewed previous commentary on the subject and concluded that the consensus in 2017 was that machine-based testimony did not implicate the Confrontation Clause.[313] She reviewed the basis for that opinion—the Supreme Court's findings that the right exists to confront "witnesses" and thus, applies to testimonial hearsay (necessarily of a person)—before ultimately concluding that, if the concerns of the framers over trial by ex parte statements meant anything, then machine-based testimony should be covered by the right as they "implicate many of the same dignitary and accuracy concerns."[314] Therefore, the right would be applicable to AI-based testimony to the same extent it applies to other, machine-assisted technology like the blood test result that required cross-examination of the lab technician who did the test in *Bullcoming v. New Mexico*.[315] In a later article, Roth outlined how that right might lead to procedural discovery rights or a broader reliability assessment, before suggesting, as she did in the original *Machine Testimony* article, that the issue merits further discussion in detail.[316] Roth is not alone in suggesting the confrontation right, when applied to the issue of machine testimony or AI-based evidence, might lead to additional safeguards or a reconceptualization of the right.[317]

Yet while that issue remained an academic debate, more recent case law has largely vindicated the position that confrontation should and does apply to AI-based testimony. In 2022, the New York Court of Appeals decided *People v. Wakefield*.[318] The case involved the admission of DNA analysis performed by TrueAllele, which the defendant challenged for both reliability and confrontation reasons.[319] While the reliability issue will be discussed below in Section IV.C, the court rejected the confrontation argument, finding, like in

312. Roth, *supra* note 5, at 2040–48 (discussing the interaction between the Confrontation Clause and AI-based evidence).

313. *Id.* at 2039–40, 2040 n.356.

314. *Id.* at 2042.

315. 564 U.S. 647, 652 (2011).

316. Andrea Roth, *What Machines Can Teach Us About "Confrontation"*, 60 DUQ. L. REV. 210, 220–21 (2022); Roth, *supra* note 5, at 2040 (stating that the confrontation issue "deserves Article-length treatment"). We agree that the issue merits detailed review and discuss that as an area for future research *infra* Section V.B.

317. *See* Cheng & Nunn, *supra* note 294, at 1113–19; Murphy, *supra* note 292, at 657–61.

318. 195 N.E.3d 19, 32 (N.Y. 2022).

319. *Id.* at 23–24, 26 (discussing, separately, the reliability and confrontation challenges).

Bullcoming, that the lab technician is not just a conduit for the machine results but also a witness, so the fact that the lab tech and the programmer testified and were subject to cross would satisfy the right.[320] The *Wakefield* court suggests that other courts would also agree with Roth, that confrontation does apply to a wide variety of machine testimony including AI-based evidence, even if the contours of that coverage will be specified at a later date.[321]

A judge applying the Confrontation Clause, as they should, to expert systems will likely reach the same conclusion as in *Wakefield*—that the right can be satisfied with the testimony of the appropriate expert.[322] As in *Bullcoming*, the expert is not a "mere scrivener" of the result but rather independently attests to the proper test parameters, and so in explaining the test management should provide sufficient basis, along with the testimony establishing the reliability necessary under Rule 702, discussed above, to satisfy the Sixth Amendment guarantee.[323] It seems likely that many cases will require both the testimony of the lab technician, to satisfy the *Bullcoming* confrontation right for the specific test, and the programmer or an AI expert, to testify to the reliability of the AI algorithms both in general and as applied.[324] This is exactly how the prosecution presented the evidence in *Wakefield*, meeting approval of New York's highest court.[325]

This is not a departure from current doctrine on analogous devices or systems. For example, drug cases routinely permit the lab analysis of the substance in question, often using devices such as a mass spectrometer to identify the substance and its potency.[326] It is admissible because, when challenged, an expert can explain the methodology a mass spectrometer uses to assess

320. *Id.* at 30–32.

321. *Id.* at 31–32 (analyzing DNA testing that had "some measure of 'artificial intelligence'" under Confrontation Clause doctrines); *see infra* Section V.B (discussing the extent to which confrontation covers machine or AI-based testimony as well as the scope of protection).

322. *Wakefield*, 195 N.E.3d at 31–32 (holding that testimony from the analyst who performed the test and a doctor who understood the "parameters and methodology" of the software satisfied the Confrontation Clause).

323. *Id.* at 30–32 (quoting Bullcoming v. New Mexico, 564 U.S. 647, 659 (2011)).

324. *Id.* at 31–32 (explaining that testimony from both the lab analyst and an expert who fully understands the software satisfies the Confrontation Clause).

325. *Id.*

326. FAIGMAN ET AL., *supra* note 8, § 40:1 ("Starting in the early 1980s, tests for drugs became commonplace in the criminal justice system . . . ."); *id.* § 40:17 ("Drug testing has achieved widespread general acceptance as a reliable scientific technique in an array of substantive areas. There simply are no cases challenging properly performed drug tests using sophisticated techniques such as gas chromatography/mass spectrometry.").

chemical compounds[327] and can also be cross-examined by the defendant (in a criminal case),[328] meeting reliability and confrontation standards.

Expert systems using rule- or case-based analysis meet the *Daubert* gatekeeping test and satisfy confrontation rights, and therefore should be admissible at civil or criminal trials, unless the evidence in a specific case fails to connect the dots between method and conclusion.

B. Application to Machine Learning and Neural Networks

If expert testimony will, in general, suffice to demonstrate the reliability of case- and rule-based expert systems, the same considerations will lead to the exclusion of all machine learning- and neural network-based AI testimony. Machine learning can be defined as "programing computers to optimize a performance criterion using example data or past experience."[329] To accomplish that task, the system will begin with a model with an initial set of parameters and then use a learning algorithm to review the data and optimize those parameters.[330] Neural networks, as a type of machine learning, involve the assessment of a problem through the interaction of perceptrons, acting like brain neurons, leading to an ultimate conclusion.[331] The process can be linear, but often involves hidden layers of perceptrons, which can involve "successive hidden layers correspond[ing] to more abstract representations until we get to the [final] output layer[.]"[332] What both methodologies have in common, as reviewed in Part III, is that they are, at a fundamental level, neither explainable nor transparent, in that the outcomes of the system lack a clear connection between the input and the output, and thus may have validity but cannot be assessed for reliability.[333]

When assessing machine learning or neural networks for reliability, *Daubert* requires an assessment not of the result alone but instead the methodology that produces the result: "The focus, of course, must be solely on

327. *Id.* § 40:2 ("The concepts underlying most chemical drug tests are well developed. The underlying principles supporting techniques such as gas chromatography/mass spectrometry are rarely challenged.").

328. *Id.* § 40:15 (discussing the applicability of the Confrontation Clause to lab testing).

329. ALPAYDIN, *supra* note 96, at 3.

330. *Id.*

331. *Id.* at 267–77.

332. *Id.* at 307–08.

333. *See supra* Section III.D.2 (reviewing the machine-learning methodologies); *see also* ALPAYDIN, *supra* note 96, at 307–08 (explaining how neural networks become more abstract with each layer).

principles and methodology, not on the conclusions that they generate."[334] No matter how attuned a machine-learning system is to the data, at a fundamental level, it will frequently lack a clear connection of the input to the outcomes.[335] For a neural network, this can be explained by the involvement of a hidden layer of perceptrons, or even multiple hidden layers in a "deep neural network."[336] The same is true for machine learning, as the association of any individual factor to the end result will necessarily remain opaque.[337] A judge being offered a neural network-based assessment of a particular piece of evidence may, under certain circumstances, be shown that the network has concluded there is a "match," and thus, the evidence is probative of guilt, but what the judge will be unable to tell, in this example and in every example of machine learning, is how the AI reached the conclusion.[338] On a fundamental level, admission of the evidence would equate to acceptance of results without their methodological foundation in science being confirmed; this is exactly what *Daubert* forbids.[339]

Of course, in the subset of cases in which AI evidence is being offered against a criminal defendant, the Confrontation Clause should affect admissibility as well. Our analysis begins with a comparison of the analysis for an expert system, discussed above in Section VI.A. In *Wakefield*, the testimony of both the lab technician, who discussed the parameters of the specific test, and the AI expert, who attested to the reliability of the system in general and as applied, satisfied the confrontation right.[340] But when dealing with machine learning- or neural-network-based testimony, the proponent is by definition

334. Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 595 (1993).

335. ALPAYDIN, *supra* note 96, at 307–08 (describing, for example, how neural networks become more abstract with each layer). Contemporary researchers in AI systems are attempting to develop methodologies for peering into the closed-boxes of those systems. *See, e.g.*, Hansen & Rieger, *supra* note 124, at 46 (discussing the use of heat maps in assessing why a particular result was obtained). The utilization of such systems in assessing AI expert evidence poses the same basic problems as AI expert systems themselves. *Id.* at 5 (describing transparency challenges facing AI).

336. ALPAYDIN, *supra* note 96, at 307–08; *see also* Samek & Müller, *supra* note 95, at 6 (discussing the complexity of deep learning networks); ALPAYDIN, *supra* note 96, at 307–09.

337. *See, e.g.*, ROTHMAN, *supra* note 218, at 1 (noting that the Markov decision process is memoryless and applies random decisions.); Weller, *supra* note 125, at 26 (discussing the difficulties in interpreting particular decisions and predictions).

338. *See* ALPAYDIN, *supra* note 96, at 307–08 (explaining how hidden layers of neural networks result in outputs that lack clear connections to inputs); *see also supra* text accompanying notes 335–337 (discussing lack of connection between input and output in machine learning).

339. *Daubert*, 509 U.S. at 595 ("The focus, of course, must be solely on the principles and methodology, not on the conclusions they generate.").

340. People v. Wakefield, 195 N.E.3d 19, 31–32 (N.Y. 2022).

unable to provide the same foundation for admission.[341] A lab technician could, in a machine-learning or neural-network case, probably attest to the fact that the specific test was done according to parameters, but there can never be a witness who can trace the general or specific reliability of the system as they lack both explainability or transparency.[342]

Confrontation is a right that exists not only to ensure outcomes that are accurate, although that is one of its purposes,[343] but also in a principle of procedural fairness to the defendant—to face the accuser and judge demeanor.[344] Without any ability to determine the methodology used in decision-making, a criminal defendant is left with an ex parte accusation presented with a technological façade.[345]

Cases addressing AI-based testimony in recent years suggest this conclusion as well. Roth cites Judge Goodwin Liu, in his 2012 dissent in *People v. Lopez*, for the proposition that admitting AI without delving into the reliability is a recreation of the abusive civil law methodologies that resulted in the Confrontation Clause in the Sixth Amendment.[346] In the *Wakefield* decision in 2022, the court did not reach the same conclusion solely because there was testimony explaining the methodology of decision-making; without foundation, the AI evidence would have been excluded.[347]

Thus, although the confrontation right does cover machine testimony, as Judge DiFiore found in *Wakefield* and Judge Liu proposed in *Lopez*, the right cannot be satisfied by opaque AI methodology.[348] Machine learning- and neural network-based evidence therefore should always fail confrontation analysis.[349]

---

341. *See* ALPAYDIN, *supra* note 96, at 307–08 (explaining how in neural networks the input and output are not connected); *see also supra* Section III.C (discussing various weaknesses in machine learning and neural network AI, including that results cannot be explained).

342. *See* Weller, *supra* note 125, at 28 (noting that a lack of transparency and explainability is a significant challenge in AI); *see also supra* Section III.C (describing the lack of transparency in certain AI formats).

343. Maryland v. Craig, 497 U.S. 836, 845–46 (1990); *see also* Roth, *supra* note 5, at 2048.

344. California v. Green, 399 U.S. 149, 157–58 (1970); *see also Wakefield*, 195 N.E.3d at 31.

345. Roth, *supra* note 5, at 2043.

346. *Id.* at 2044 (citing People v. Lopez, 286 P.3d 469, 494 (Cal. 2012) (Liu, J., dissenting)).

347. *Wakefield*, 195 N.E.3d at 31–32.

348. *See id.* (discussing the Confrontation Clause as applied to machine testimony); *Lopez*, 286 P.3d at 494 (applying Confrontation Clause analysis to machine testimony).

349. *See supra* Sections III.C–III.D (discussing transparency and explainability, and whether AI systems meet those requirements). To date, we are not aware of any state or federal case that addresses

In summary, machine-learning and neural-network AI evidence, by their very natures, fail both the *Daubert* gatekeeping test and Sixth Amendment confrontation, rendering this kind of AI evidence inadmissible at civil and criminal trials. Even so, such methodologies can be used, under appropriate regulation,[350] for law enforcement to generate leads or otherwise assess complex phenomena, but such usage must—under the current rules—stop at the courthouse door.

C. The Intermediate Case–Bayes and Fuzzy Logic Expert Systems

By applying the rules for gatekeeping and, when applicable, the confrontation right to AI-based evidence, the admission should be based on the computational design, and because of the opacity of the decision-making process, neural networks and machine learning are inadmissible while expert systems can be permitted with the proper basis.[351] The final example of AI-based evidence, the collection of both Bayesian and fuzzy-logic expert systems, provides an intermediate example stretching the current doctrine to its limit.[352] Ultimately, we conclude that AI evidence based on Bayes or fuzzy logic can and should be admitted as reliable and in accord with confrontation rights, although we suggest caution for judges in addressing these considerations.

A quick review of the computational methodologies leads us to this conclusion. In Bayesian systems, the system will evaluate different rules but assign probability (or certainty) values to them for a more complex understanding of the interaction of the variables.[353] In a fuzzy expert system, the knowledge base is composed of a collection of fuzzy rules, fuzzy membership functions, and associated fuzzy information that represents the (imprecise) expertise and domain knowledge of the system.[354] Both systems rely on complex statistical, logical, and mathematical reasoning to reach an ultimate

machine learning- or neural network-based evidence. We would expect this to change soon as the technology advances.

350. *See infra* Section V.A (exploring the current state of regulation and the need for more AI regulation in the United States).

351. *See supra* Sections III.C–III.D (explaining the opacity of various systems).

352. *See supra* Sections III.D.1.c–III.D.1.d (discussing these computation methodologies in further detail).

353. *See supra* text accompanying notes 175–184.

354. *See* Abraham, *supra* note 136, at 912–13; *see also supra* text accompanying notes 185–197.

conclusion.[355]

When applying the reliability standard of *Daubert* to these rules, it is imperative that the court receive evidence allowing the judge to review the computation that leads to a specific conclusion.[356] Only this evidence will suffice for reliability standards to be met because *Daubert* explicitly reminds us that the "focus, of course, must be solely on principles and methodology, not on the conclusions that they generate."[357] When evaluating the computational methodology, if the evidence leaves "too great an analytical gap" then the conclusions are unreliable.[358]

Bayesian systems should be able to meet the requirement of reliability if the proper testimony can be provided by an expert in the AI design. Since the computational methodology involves a series of conditional probabilities using a variety of statistical models, an expert should be able to "unpack" the computational methodology of the system in question and explain its overall architecture in addition to the inferences made in the specific case.[359] The same is true for fuzzy logic, which create fuzzy rules to capture the membership relationship between variables, leading to an ultimate conclusion.[360] So, if the proponent can support the inferences used by the AI system in general and in the specific case, then the system is theoretically identical to the rule- and case-based systems and should be admitted.

So why the hesitancy? Our concern about admission of Bayesian and fuzzy logic systems is not as much based on the legal constraint of reliability—because it can be met—but instead on a practical limitation of the jury to understand these complex systems. Juries are not well-known for their understanding of statistical reasoning; in fact, many studies question their ability to handle complex expert testimony.[361] If juries cannot understand the statistical

355. *See* Abraham, *supra* note 136, at 912–13 (explaining fuzzy expert systems); *see also* ALPAYDIN, *supra* note 96, at 49–64 (explaining Bayesian theory); *see also supra* Sections III.D.1.c–III.D.1.d (discussing Bayesian theory and fuzzy expert systems in detail).

356. Roth, *supra* note 5, at 2047–48.

357. Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 595 (1993).

358. Gen. Elec. Co. v. Joiner, 522 U.S. 136, 146 (1997).

359. *See* ALPAYDIN, *supra* note 96, at 49–64 (explaining Bayesian systems); *see also* Section III.D.1.c (discussing the computation method used by Bayesian systems).

360. *See supra* text accompanying notes 185–197.

361. For a detailed examination of the literature assessing juror management of complex evidence, see Andrew W. Jurs, *Expert Prevalence, Persuasion, and Price: What Trial Participants Really Think About Experts*, 91 IND. L.J. 353, 360–63 (2016), and Neil Vidmar & Shari Seidman Diamond, *Juries and Expert Evidence*, 66 BROOK. L. REV. 1121, 1140–49 (2001). *See generally* FAIGMAN ET AL., *supra* note 8, §§ 3:15–3:25 (reviewing the issue of jury management of complex expert evidence).

reasoning behind AI systems, the court must recognize that fact and, to some extent, consider it as one factor among many in the admission of the evidence as reliable. For example, imagine a polygraph machine that has a clearly defined but complex statistical methodology that leads to reliable results. Basic doctrine would suggest it would be reliable enough to admit under *Daubert*.[362] Yet the admissibility decision must also take into account the jury effect, and so judges may, even if the machine is reliable, decide it infringes too much on the jury role.[363] The same can be true of complex statistical modeling using these AI expert systems—even if the system is explainable by an expert, if the jury doesn't understand the basics, it suggests too much deference to the machine and, like with neural networks, raises the issue of "trial by machine."[364] We do not, however, suggest this is enough reason to presumptively exclude Bayesian and fuzzy logic expert systems, but instead we offer a suggestion of extreme caution considering the risks.

Recent case law suggests the reliability hurdle can be cleared when using Bayesian systems. One recent Sixth Circuit case, *United States v. Gissantaner*, addressed a reliability challenge to DNA evidence based on a system called STRMix.[365] STRMix uses statistical modeling, specifically a Markov Chain Monte Carlo method, to create a likelihood ratio between two competing hypotheses which then uses a national database for a final calculation of probability.[366] When challenged, the Sixth Circuit examined the STRMix analysis using the standard *Daubert* factors, and concluded that it was testable, subject to peer review, with a low error rate, and general acceptance within the lab community, thus overruling exclusion from the district

362. *See Daubert*, 509 U.S. at 590–93 (holding that expert testimony must be backed by scientific knowledge and an explanation of the results to be considered reliable).

363. Polygraph examinations are a typical example of the assessment of jury effect overriding reliability considerations, resulting in exclusion of the evidence even after the *Daubert* standard. FAIGMAN ET AL., *supra* note 8, § 38:5 ("Polygraph tests present a particularly appropriate example of the importance of Rule 403 in managing scientific evidence."); Roth, *supra* note 18, at 1257 (discussing prosecutors' concerns that polygraphs interfere with jury credibility determinations); *id.* at 1293 ("[C]ourts have almost universally rejected so-called lie-detection evidence of credibility at trial on grounds that it is unreliable or that, even if reliable, it would usurp the jury's credibility-determining role.").

364. Roth, *supra* note 18, at 1293; *see also* Moriarty, *supra* note 300, at 245 (discussing whether the judge and jury can handle these reliability questions).

365. 990 F.3d 457, 460 (6th Cir. 2021).

366. Whittley v. State, No. 05-21-00534-CR, 2022 WL 3645589, at *6 (Tex. Ct. App. Aug. 24, 2022).

court.[367] The decision is the first court of appeals analysis of the STRMix approach, although the court did cite several district court opinions in accord with their finding.[368] At the state level, a 2022 Texas Court of Appeals case, *Whittley v. State*, reached a similar conclusion regarding STRMix.[369] After a review of the foundation provided to the trial judge, the court found that "the underlying scientific theory and technique can be clearly explained to the court," and that the expert in the case did so.[370] Under these circumstances, as with *Gissantaner*, the reliability had been demonstrated, making admission to the jury proper.[371] Cases addressing the issue of reliability under *Frye*-based state rules have reached similar conclusions.[372]

Yet, in addition to reliability, certain cases will also raise confrontation concerns. As discussed above, the Confrontation Clause does apply to AI-based evidence, although few courts have yet to address the implications.[373] The *Gissantaner* and *Whittley* cases, for example, address reliability alone and few cases address confrontation in the context of AI evidence.[374] The sole case addressing confrontation with STRMix is *Wakefield*, which discussed both reliability and confrontation challenges.[375] The New York court in *Wakefield* applied a *Frye*-based state reliability standard, finding that the statistical modeling did meet the "general acceptance" required to admit.[376] The court then addressed confrontation, finding that confrontation did apply to AI-based

367. *Gissantaner*, 990 F.3d at 463–66.

368. *Id.* at 466 (citing six prior admissions in federal district court and several other non-federal decisions).

369. *Whittley*, 2022 WL 3645589, at *1.

370. *Id.* at *7.

371. *Id.* at *8.

372. *See, e.g.*, People v. Wakefield, 195 N.E.3d 19 (N.Y. 2022); People v. Davis, 290 Cal. Rptr. 3d 661, 686 (Cal. Ct. App. 2022). *Frye* states are those jurisdictions that admit expert testimony when an issue involves specialized subject matter outside of "common knowledge." Frye v. United States, 293 F. 1013, 1014 (D.C. Cir. 1923) ("When the question involved does not lie within the range of common experience or common knowledge, but requires special experience or special knowledge, then the opinions of witnesses skilled in that particular science, art, or trade to which the question relates are admissible in evidence." (quotation omitted)).

373. *See supra* text accompanying notes 312–324 (discussing the applicability of the Confrontation Clause to AI); *see also* Sites, *supra* note 131, at 548–49 (noting that many courts have concluded machines are outside the scope of the Confrontation Clause).

374. United States v. Gissantaner, 990 F.3d 457, 463–69 (6th Cir. 2021) (addressing reliability); *Whittley*, 2022 WL 3645589, at *4–8 (discussing reliability); Roth, *supra* note 316, at 210–11 (explaining courts' narrow construction of the Confrontation Clause and its relationship to AI).

375. *Wakefield*, 195 N.E.3d at 21 (explaining the appeal deals with both reliability and confrontation challenges).

376. *Id.* at 27–30.

evidence, as reviewed earlier, and that the testimony of the lab technician and the software designer were sufficient to meet the requirements of the Confrontation Clause.[377] The result here is instructive and likely to be the result in many of the challenges to these types of expert systems because, to the extent the foundation has been laid to establish reliability, both in general and in the specific case, the cross-examination of those witnesses means the right to confront has been granted.[378] The only hesitation from *Wakefield* is based on its unusual substantive basis because, as a *Frye* state, the foundation necessary to establish admission through general acceptance may not (always) be enough to provide full confrontation rights.[379] Yet, it seems clear that when handling reliability under the broader *Daubert* standard, as in *Gissantaner*, the confrontation right will be met.[380] Perhaps this explains why so many of the *Daubert* cases lack a parallel confrontation appeal argument.[381]

When applying the current legal standards to Bayes and fuzzy logic expert systems, we conclude that they are, when accompanied by the appropriate foundation testimony, likely to be admitted as reliable and to satisfy confrontation standards of the Sixth Amendment. Even if so, their reliance on complex statistical reasoning and the confusion likely to be generated from the foundational testimony establishing reliability should be considered when evaluating admission to ensure that the jury is not merely a conduit for a "trial by machine."[382] In that regard, the AI-based evidence is not dissimilar from other complex evidence judges routinely handle.[383]

Having reviewed the rules of reliability and the Confrontation Clause

377. *Id.* at 30–32.

378. *See* Sites, *supra* note 131, at 552–59 (discussing various cases and how courts have deemed the cross-examination of technicians as satisfying the Confrontation Clause); *see also* California v. Green, 399 U.S. 149, 153 (1970) (explaining that the right to cross-examine is satisfied if it occurs prior to or during trial).

379. *Wakefield*, 195 N.E.3d at 28–32 (discussing the general acceptance standard and what testimony satisfies it).

380. United States v. Gissantaner, 990 F.3d 457, 463–69 (6th Cir. 2021) (discussing reliability under the *Daubert* standard).

381. *See* David H. Kaye & Jennifer L. Mnookin, *Confronting Science: Expert Evidence and the Confrontation Clause*, 2012 SUP. CT. REV. 99, 100–04 (2013) (analyzing *Daubert* cases and the Confrontation Clause).

382. *See generally* Roth, *supra* note 18 (exploring the historical rise of machines and how it impacts the criminal justice system).

383. *See* Thomas D. Albright, *A Scientist's Take on Scientific Evidence in the Courtroom*, PNAS (Oct. 2, 2023), https://www.pnas.org/doi/10.1073/pnas.2301839120 (discussing scientific evidence that is complex and may require expert testimony); *see also* text accompanying notes 361–364 (discussing polygraph evidence and its complexities).

issue, we conclude that expert systems should be admissible under both, so long as the reliability can be established for the computational methodology through appropriate expert testimony. On the other hand, machine learning and neural network AI applications should be excluded as failing reliability—because the focus of the assessment will be on the methodology not conclusions—and therefore violate the confrontation rights of the accused when used against a criminal defendant.

V. AREAS FOR LEGAL DEVELOPMENT AND FUTURE ANALYSIS

If anything remains predictable about the development of AI, it is the rapid rate of development that has, in a year or two, turned a conversation about theoretical possibilities into one of pressing concerns about current implementation of technology.[384] So, to some extent this Yogi Berra-attributed saying seems appropriate: "Prediction is difficult, especially about the future."[385] Yet, based on the current analysis of both the state of computational methodologies in AI and also the state of the law, we can offer several suggestions for policy reform in the area. Afterward, we suggest several areas for future work which should begin to answer our most pressing questions about the applicability of the law to the field. In a field of this magnitude, and with the pace of technological development, however, these are more than likely just the start.

A. Policy Prescriptions for the Field

Our analysis concludes that some AI evidence meets the requirements of reliability and, when applicable, confrontation necessary for admission at trial, while some does not, largely based on an assessment of the computational methodology underlying the AI analysis. We also identified potential pitfalls of admission of AI-based evidence, particularly the complexity of explanation of some analytical methods and the limits of confrontation to

384. Grimm et al., *supra* note 5, at 12 ("AI algorithms are no longer the stuff of science fiction or the imagination of high-tech brainiacs. They are being used right now, in countless software applications, and in increasingly expansive ways . . . ."); Kathy Baxter & Yoav Schlesinger, *Managing the Risks of Generative AI*, HARV. BUS. REV. (June 6, 2023), https://hbr.org/2023/06/managing-the-risks-of-generative-ai (discussing concerns about the implementation of AI).

385. *It's Difficult to Make Predictions, Especially About the Future*, QUOTE INVESTIGATOR, https://quoteinvestigator.com/2013/10/20/no-predict/ (last visited Jan. 31, 2024). The origin of the quote has been the subject of debate. *Id.*

address AI issues. These basic conclusions lead directly to several policy proposals that would address weaknesses or lack of clarity in the current system.

First, it is very clear that AI-based evidence, even when reliable and subject to cross, has a significant potential to confuse or overwhelm the jury at trial.[386] It is not the only evidence to have this potential, as experts can be admitted to address a wide variety of complex fields (so long as the proper foundation has been provided).[387] Judges must recognize, however, the longstanding concern that jurors would allow expertise to overcome their own assessment of the evidence and, instead, blindly agree with the infallible result of advanced computing AI.[388] Our concern for this is particularly acute when the computational methodology in question is reliable, but statistically complex such as in Bayesian or fuzzy logic expert systems.[389] While considerable effort will be necessary for judges to assess these AI systems for reliability, as discussed in Section IV.C, even if they are admitted, additional caution is appropriate. So, we propose that jury instruction commissions, judicial panels, circuits, or other parties who draft instructions consider adopting a cautionary instruction for AI-based evidence. Such an instruction should remind jurors that the AI-based evidence is solely one part of the analysis, that the opinions generated are only as good as the underlying analytical methodology, and ultimately the decision to accept or reject the evidence, in whole or in part, should remain with the jury alone.[390] Similar instructions have been offered to courts in DNA cases, although not always accepted, and they can offer jurors guidance in an area of particular complexity.[391]

386. Grimm, et al., *supra* note 5, at 88.

387. *See* Daubert v. Merrell Dow Pharm., Inc., 590 U.S. 579, 597 (1993) (holding that scientific evidence is admissible under federal evidence rules if the judge ensures a "reliable foundation" is created by the expert testifying).

388. *See* Anne E. Boustead & Matthew B. Kugler, *Juror Interpretations of Metadata and Content Information: Implications for the Going Dark Debate*, 9 J. CYBERSECURITY 1, 2 (2023) ("[J]urors might be predisposed to trust . . . technically derived evidence . . . .").

389. *See supra* Sections III.D.1.c—III.D.1.d (discussing Bayesian and fuzzy logic systems); *see also* Piero Baraldi et al., *Comparing the Treatment of Uncertainty in Bayesian Networks and Fuzzy Expert Systems Used for a Human Reliability Analysis Application*, 138 RELIABILITY ENGINEERING & SYS. SAFETY, 176, 176 (2015) (stating that Bayesian and fuzzy logic systems are complex but understandable).

390. *See* Vidmar & Diamond, *supra* note 361, at 1128–31 (discussing the issue of jury instructions for expert testimony in general); *see also* KEVIN F. O'MALLEY ET AL., FEDERAL JURY PRACTICE AND INSTRUCTIONS §14:01 (Thomson West, 6th ed. 2022) (collecting examples of expert witness instructions used in federal court for expert testimony).

391. *See generally* Pooja Chaudhuri, *A Right to Rational Juries? How Jury Instructions Create the*

Second, as we have concluded that the admission of AI-based evidence depends largely on the computational methodology underlying the analysis, we propose clarity for the methodologies to be an inherent part of any AI design. If the judicial assessment of AI technology is based on how the analysis occurs, then the AI system must have the transparency to evaluate the complete assessment of that method.[392] One could imagine that a particular conclusion could be the product of mixtures of various AI computational methodologies, for example a rule-based expert system supplemented by a neural network.[393] In these cases, regulation and voluntary compliance should combine to establish a norm of openness permitting the level of assessment required for gatekeeping.

Several commentators in the field have suggested that this openness and transparency must, by its very nature, lead to an enhanced right of discovery for any AI-based evidence. Erin Murphy, in a 2014 article on confrontation rights, suggested that openness would require new rules "amplifying the right to discovery to include things such as an analyst's historical error reports, proficiency test results, and other performance evaluations."[394] She also questions whether the confrontation right could lead to a right to enhanced discovery, a so called "collective confrontation right to transparency and accountability standards in forensic analysis."[395] Andrea Roth addressed the same issue in *Machine Testimony* in 2017, stating that "machine sources should not be given an absolute pass under the [Confrontation] Clause," and later suggesting that this could include access to the source code in addition to expert testimony from the programmer.[396] She continued this analysis in a later paper in 2022, suggesting again that broader discovery, prior machine

"Bionic Juror" in Criminal Proceedings Involving DNA Match Evidence, 105 CALIF. L. REV. 1807 (2017) (explaining the importance of instructions in DNA cases). Chaudhuri offers model language for the instructions for DNA cases. *Id.* at 1850–52.

392. *See supra* Section III.D.4 (discussing the requirements of transparency); *see also* Grimm et al., *supra* note 5, at 60 (noting a solution is to require transparency of how the system works and how it reached a decision).

393. *See* Ben Dickson, *What Happens When You Combine Neural Networks and Rule-Based AI?*, TECHTALKS (June 5, 2019), https://bdtechtalks.com/2019/06/05/mit-ibm-hybrid-ai/#:~:text=In%20the%20hybrid%20AI%20model,things%20with%20much%20less%20data (explaining that in a hybrid AI system, the rule-based system "takes advantage of the neural networks' ability to process and analyze unstructured data," while the neural network "benefits from the reasoning power of the rule-based AI system.").

394. Murphy, *supra* note 292, at 659.

395. *Id.* at 659–60.

396. Roth, *supra* note 5, at 2047–48, 2050.

output in other cases, or independent testing could be part of the right.[397] Edward Cheng and Alex Nunn brought a broad process approach to confrontation assessment in their 2019 work, *Beyond the Witness*, calling the confrontation right's obsession with witness testimony "antiquated and counterproductive."[398] They then proposed, as Murphy and Roth did, that a broader confrontation right would involve assessment of the underlying process, including inquiry into "the processes' accuracy, transparency, and objectivity."[399] Our call for additional transparency is consistent with these proposals, although one could imagine an intermediate step of discovery and process at the sub-constitutional level involving modification of the rules of procedure to address the same considerations. Clearly the limits of such processes will continue to be debated, if not the actual need for transparency and discovery, for AI-based evidence.

Along similar lines, government needs to be more assertive and regulate AI methodologies, implementation, and use before it becomes ubiquitous. The European Union has so far been ahead of the United States in this area, as the European Parliament has already approved a ban on AI use for facial recognition and predictive policing.[400] The AI Act, approved by the Parliament on June 14, 2023, also includes disclosure and transparency standards for AI-generated content.[401]

Meanwhile, in the United States, tech giants in the AI field have repeatedly testified to Congress about the need for enhanced regulation, so far without any significant progress; Microsoft's president specifically warned that the "[g]overnment needs to move faster."[402] Senate Majority Leader Chuck Schumer proposed a broad legislative framework, called SAFE Innovation in

397. Roth, *supra* note 316, at 222.

398. Cheng & Nunn, *supra* note 294, at 1078.

399. *Id.* at 1122.

400. Emma Woollacott, *Draft AI Act Passes, Banning Police Facial Recognition*, FORBES (June 15, 2023, 4:57AM), https://www.forbes.com/sites/emmawoollacott/2023/06/15/draft-ai-act-passes-banning-police-facial-recognition/?sh=4bc7ff3b4965. For the official text of the AI Act, see *Artificial Intelligence Act*, COM (2021) 0206 (June 14, 2023), https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.pdf.

401. Woollacott, *supra* note 400.

402. David McCabe, *Microsoft Calls for A.I. Rules to Minimize the Technology's Risks*, N.Y. TIMES (May 25, 2023), https://www.nytimes.com/2023/05/25/technology/microsoft-ai-rules-regulation.html; *see also* Cecelia Kang, *How Sam Altman Stormed U.S. Congress to Set the AI Agenda*, N.Y. TIMES (June 7, 2023), https://www.nytimes.com/2023/06/07/technology/sam-altman-ai-regulations.html.

the AI Age in the summer of 2023,[403] emphasizing that the legislative process is only beginning, and stated the timeline for advancement is "not going to be days or weeks, but it's not going to be years. Months would be the proper timeline."[404]

Since national reform seems stalled, two smaller changes are notable.[405] Even without comprehensive legislation to guide them, states and localities are adopting legislation or rules to limit their own use of AI tools.[406] States have adopted limits on AI use in insurance markets and hiring decisions,[407] and police departments in New York City and Detroit have issued formal policies limiting use of facial recognition technology.[408] Such efforts are to be applauded, but are both inadequate in scope and way behind implementation of current AI.[409] In fact, they remain so inadequate that the private sector has, in the interim, adopted safety limits without governmental involvement.[410] These regulatory efforts need additional urgency and, to return to AI-based

403. Chuck Schumer, *SAFE Innovation Framework*, SENATE DEMOCRATS, https://www.democrats.senate.gov/imo/media/doc/schumer_ai_framework.pdf (last visited Feb. 27, 2024); *see also* Interview with Senator Chuck Schumer, Senate Majority Leader (June 21, 2023) [hereinafter Schumer Interview], https://csis-website-prod.s3.amazonaws.com/s3fs-public/2023-06/230621_Schumer_SAFE_Innovation.pdf.

404. Schumer Interview, *supra* note 403; *see also* Cecelia Kang, *In U.S., Regulating A.I. Is in Its 'Early Days'*, N.Y. TIMES (July 21, 2023), https://www.nytimes.com/2023/07/21/technology/ai-united-states-regulation.html.

405. *See* Brendan Bordelon, *On AI, the Government Gets Ready to Throw Its Weight Around*, POLITICO (May 16, 2023, 5:26 PM), https://www.politico.com/news/2023/05/16/the-government-plots-its-ai-approach-00097262 (noting that with federal legislation stalled, legislating specifically on federal agencies may be the best way to regulate); Lawrence Norden & Benjamin Lerude, *States Take the Lead on Regulating Artificial Intelligence*, BRENNAN CTR. FOR JUST. (Nov. 6, 2023), https://www.brennancenter.org/our-work/research-reports/states-take-lead-regulating-artificial-intelligence (discussing state legislation on AI).

406. *See, e.g.*, Norden & Lerude, *supra* note 405 (exploring state legislation on AI); *Legislation Related to Artificial Intelligence*, NCSL (Jan. 31, 2023), https://www.ncsl.org/technology-and-communication/legislation-related-to-artificial-intelligence (tracking state legislation on AI).

407. *See US State-by-State AI Legislation Snapshot*, *supra* note 55.

408. N.Y. POLICE DEP'T, FACIAL RECOGNITION: IMPACT AND USE POLICY 2 (2021), https://www.nyc.gov/assets/nypd/downloads/pdf/public_information/post-final/facial-recognition-nypd-impact-and-use-policy_4.9.21_final.pdf; DETROIT POLICE DEP'T MANUAL §307.5 (2019), https://detroitmi.gov/sites/detroitmi.localhost/files/2020-10/307.5%20Facial%20Recognition.pdf.

409. Adam Satariano & Cecilia Kang, *How Nations are Losing a Global Race to Tackle A.I.'s Harms*, N.Y. Times (Dec. 6, 2023), https://www.nytimes.com/2023/12/06/technology/ai-regulation-policies.html#:~:text=a%20funda ("A.I. systems are advancing so rapidly and unpredictably that lawmakers and regulators can't keep pace.").

410. Kang, *supra* note 404 (noting that seven companies announced voluntary AI limits at the White House).

evidence, should require additional transparency and discovery adequate to manage the gatekeeping necessary for admission in court.

Policy advancement in the field can include mitigation of the jury effect through effective instructions, expanded discovery and transparency for AI, and effective regulation.

B. Most Pressing Areas for Future Analysis

Even when addressing the issue of admission of AI-based evidence, this paper has necessarily been limited to addressing certain pressing topics. Yet, for a topic of this magnitude, additional considerations merit assessment and we intend to examine them in future work.

The primary issue that remains undecided, even within the considerations we did address, is the scope and limits of the confrontation right for AI-based evidence. In Section IV.A, we discussed the debate on whether confrontation applies to machine testimony, including AI-based evidence,[411] and later, in Section V.A., we reviewed assessments from scholars like Roth, Murphy, and Cheng and Nunn, each on the impact that coverage could have on procedures and disclosure obligations.[412] In 2022, the New York Court of Appeals held in *People v. Wakefield* that confrontation should apply to AI-based testimony.[413] This decision is an important first step, but clearly leaves many questions unanswered. Unanswered questions include: Will all courts agree with *Wakefield* that confrontation applies? Does it apply equally to all types of AI-based evidence? Does confrontation require additional discovery or procedural safeguards to work? And should these questions be resolved at the constitutional level or can they be more appropriately addressed through rules, statutory amendment, or other similar means?

Commentary so far has suggested answers to some of these questions, but additional debate, analysis, and case law should continue to do so.[414] The primary issue that remains undecided, but was not part of the considerations in this work, involves the material used to train an AI machine. Some types of AI, by their design, will incorporate vast amounts of information into their

411. *See supra* Section IV.A.

412. *See supra* Section V.A.

413. 195 N.E.3d 19, 31–32 (N.Y. 2022).

414. *See, e.g.*, Ronald J. Coleman & Paul F. Rothstein, *A Game of* Katso *and Mouse: Current Theories for Getting Forensic Analysis Evidence Past the Confrontation Clause*, 57 AM. CRIM. L. REV. 27, 56 (2020) (noting that these issues will only become more pervasive and the continuing importance of safeguarding constitutional rights).

learning cycles as part of the algorithm training.[415] ChatGPT, as one example, uses information on the internet to train its algorithm to be able to answer user inquiries.[416] Yet, the material used for training may not be clear to the programmer or user, leading to significant unanswered questions.[417] In the context of intellectual property, Hollywood has raised the concern that use of copyrighted material for AI training should result in compensation to the creator.[418] In the context of evidence for trial, however, expert evidence not only must be shown to be reliable, as discussed throughout this Article, but it must also be "based on sufficient facts or data" appropriate to the field.[419] Use of unknown data for AI training may run afoul of both provisions because, by definition, if the data of the training algorithm is unknown or unknowable then it may not be either sufficient or field-appropriate.[420] We know of no commentary addressing this consideration, but intend to examine it in future work.

As AI expands its influence on a variety of human systems and processes, the use in court should and must be limited to only those instances which meet the current standards for admission based on expert testimony.[421] By reviewing the different types of AI systems, we have provided a framework for assessing the admissibility of AI-based evidence based on the underlying computational methodology of the algorithm.[422] Some AI systems—like rule- or case-based expert systems—should meet the standards of reliability and confrontation necessary for admission, while neural networks and machine

415. *See supra* text accompanying notes 242–267.

416. Shreya Johri, *The Making of ChatGPT: From Data to Dialogue*, SCI. NEWS (June 6, 2023), https://sitn.hms.harvard.edu/flash/2023/the-making-of-chatgpt-from-data-to-dialogue/ ("The training dataset consisted of text collected from multiple sources on the internet, including Wikipedia articles, books, and other public webpages.").

417. *See id.* (discussing limits on ChatGPT).

418. Joseph Gordon-Levitt, *If Artificial Intelligence Uses Your Work, It Should Pay You*, WASH. POST (July 26, 2023, 6:45AM), https://www.washingtonpost.com/opinions/2023/07/26/joseph-gordon-levitt-artificial-intelligence-residuals/.

419. FED. R. EVID. 702 (explaining that expert opinions are admissible but must show the "testimony is based on sufficient facts or data."); FED. R. EVID. 703 (explaining that when the factual basis for expert testimony includes inadmissible evidence, it must demonstrate that such data is of a type that an expert in the field would reasonably rely on it in order to admit it).

420. *See* CYNTHIA CWIK ET AL., AM. ASS'N FOR THE ADVANCEMENT OF SCI., ARTIFICIAL INTELLIGENCE AND THE COURTS: MATERIALS FOR JUDGES 6–18 (2022) (discussing challenges in "evaluating the trustworthiness of AI evidence, which, in the context of court cases, means its relevance, validity, reliability, and authenticity." (emphasis omitted)).

421. *See id.* at 16–17 (explaining that expert testimony standards under Federal Rule of Evidence 702 and the *Daubert* factors should govern AI testimony admissibility).

422. *See supra* Parts III–IV (discussing AI methods and how current rules apply to them).

learning do not.[423]

Faced with a proffer of AI-based evidence at trial, we suggest, then, that judges must focus in detail on the computational methodology to ensure that it is explained to a level of detail that prevents "too great an analytic gap" between analysis and conclusion.[424]

423. *See supra* Part IV (analyzing how the current rules apply to various AI systems).

424. Gen. Elec. Co. v. Joiner, 522 U.S. 136, 146 (1997) (holding a court can decide that "there is too great of an analytical gap between the data and the opinion proffered").