Summer 2020

Highlighting the Failure of Criminal Courts to Adequately Test Machine Evidence

Francesca Laguardia

From the Legal Literature Highlighting the Failure of Criminal Courts to Adequately Test Machine Evidence

Francesca Laguardia, J.D., Ph.D.

I. Introduction

The risks modern technology poses to traditional understandings of Fourth Amendment privacy protections is a well explored topic.[1] However, recently, scholars have become aware of another risk posed by increasing technological advances to our traditional understandings of criminal law and criminal justice. Although technological advances have generally increased the accuracy of the criminal justice process,[2] their treatment at trial introduces opportunities for the jury to hear misleading or outright false information without the traditional safeguards that the rules of evidence and Confrontation Clause[3] protections have offered. Today, machines regularly provide evidence, and while that evidence may carry many of the same risks of inaccuracy or outright falsehood that witness statements carry,[4] machine evidence generally evades the scrutiny to which we subject more traditional forms of evidence.[5] But how can juries evaluate the credibility of a machine? What new protocols are necessary, and to what extent are the rules of evidence failing in regards to these new forms of evidence? This issue is only beginning to receive recognition. The following are two articles that have begun mapping the problem.

---

[1] *See*, *e.g.*, Katherine Strandburg, *Home, Home on the Web and Other Fourth Amendment Implications of Technosocial Change*, 70 Md. L. Rev. 614 (2011); Michael Rich, *Machine Learning, Automated Suspicion Algorithms, and the Fourth Amendment*, 164 U. Pa. L. Rev. 164 (2015); Lauren Fash, *Automated License Plate Readers: The Difficult Balance of Solving Crime and Protecting Individual Privacy*, 78 Md. L. Rev. Online 63 (2019).

[2] Andrea Roth, *Machine Testimony*, 126 Yale L. J. 1972, 1976 (2017).

[3] U.S. Const. amend. VI; *see also Crawford v. Washington*, 541 U.S. 36, 124 S. Ct. 1354, 158 L. Ed. 2d 177, 63 Fed. R. Evid. Serv. 1077 (2004) (holding that out-of-court statements by witnesses that are testimonial are barred, under the Confrontation Clause, unless witnesses are unavailable and defendants had prior opportunity to cross-examine witnesses, regardless of whether such statements are deemed reliable by court).

[4] Roth, *supra* note 2, at 1989–93; Brian Sites, *Machines Ascendant: Robots and the Rules of Evidence*, 3 Geo. L. Tech. Rev. 1, 21–23 (2018).

[5] Sites, *supra* note 4, at 5–6.

II. Andrea Roth, *Machine Testimony*, 126 YALE L. J. 1972 (2017).

While the problem of the increasingly routine use of technologically complex evidence has been percolating for over a dozen years,[6] the first scholar to zero in on this problem at the trial level is Andrea Roth. In her article in the *Yale Law Journal*, Roth has explored the position of machine evidence at trial, as compared to a routine witness. As Roth notes, more and more often machines offer assertions of truth, whether as to the location of an individual at time of arrest, the level of alcohol in an individual's bloodstream, or the proper interpretation of DNA panels.[7] These assertions, she argues, carry dangers similar to the hearsay dangers of human sources of evidence, which she enumerates as insincerity, ambiguity, memory loss, and misperception.[8] She refers to the respective dangers present in the case of machine testimony as "black box dangers," which, she states, are not present in all machine testimony, but when present may go unnoticed and unexamined by the jury.[9]

Roth first explains the issue of machine credibility. While some courts have suggested that machines merely convey the assertions of their programmers, she states, in truth many machines now make statements—even testimonial statements—that their programmers or the analysis using the machines could never have independently affirmed or denied.[10] While acknowledging that information from a programmer may be useful when evaluating machine conclusions, she suggests that a better comparison is to experts (with the machines as the experts) who may well rely on witnesses or other input to come to conclusions but nevertheless come to their own conclusion.[11]

But the philosophical questions of whether a machine makes independent assertions or has independent thought are less relevant to the question of fairness and accuracy at trial than the possible errors that machine sources are allowed to bring into a trial. It is to these errors that Roth turns next. She notes the possibility of falsehood by design (which might be compared to a witness who lies on the stand)—these may occur for instance when a programmer or designer purposefully designs a machine to give false or misleading

---

[6]Erin Murphy, *The New Forensics: Criminal Justice, False Certainty, and the Second Generation of Scientific Evidence* 95 CALIF. L. REV. 721 (2007); Erin Murphy, *The Mismatch Between Twenty-First Century Forensic Evidence and Our Antiquated Criminal Justice System*, 87 S. CAL. L. REV. 633 (2014).

[7]Roth, *supra* note 2, at 1971–72.

[8]Roth, *supra* note 2, at 1977.

[9]Roth, *supra* note 2, at 1977–79.

[10]Roth, *supra* note 2, at 1986–87.

[11]Roth, *supra* note 2, at 1987–88.

statements (Roth uses the example of Volkswagen's misleading emissions evaluations).[12] Additionally, machine learning allows for the possibility that machines would learn to lie in order to achieve a desired outcome. This has been achieved in some machines, although none currently used in criminal proceedings.[13] She also notes the possibility that a machine can be inarticulate or misleading due to imprecise language, for instance if insufficient information is given as to the conditions under which the machine comes to its conclusions. At what percentage of certainty does the machine state it has found an answer? What potential errors exist? What level of toleration for false positives or false negatives has the programmer used? If answers to questions such as these are not available, the machine might be considered "inarticulate," as in: it is difficult to state precisely what the machine is saying.[14] These ambiguities may also stem from human mistake in using the machine, or the degradation of the machine over time.[15]

Similarly, degradation or human mistake can create a machine that miscodes, misreads, or analyzes material incorrectly.[16] Unconscious bias by a programmer may lead to an algorithm that relies on that bias to determine what is most likely, or what is most reliable—in the case of legal analysis, such bias has led less popular precedents to all but disappear from some legal research.[17] Machine learning can lead to overfitting—assuming a relationship exists simply because a coincidental pattern has been found (as an example, Roth offers a crime analysis program that taught itself that people "who shak[e] hands three times [are] likely engaged in a drug transaction.").[18] And of course, any machine that relies on a human to input material may provide flawed results when the operator of the machine operates it or inputs material incorrectly, such as by waiting too long to run a test, or inputting the wrong sample.[19]

Not all machine evidence brings in questions of credibility, so Roth spends Section II breaking down which types of evidence might raise these problems and which most likely would not.[20] Implicit in her analysis is an analogy to hearsay determinations. Machines whose statements are not offered for the truth of the matter asserted

---

[12]Roth, *supra* note 2, at 1990.

[13]Roth, *supra* note 2, at 1992.

[14]Roth, *supra* note 2, at 1992–93.

[15]Roth, *supra* note 2, at 1993.

[16]Roth, *supra* note 2, at 1994–95, 1999.

[17]Roth, *supra* note 2, at 1995–96.

[18]Roth, *supra* note 2, at 1996–97

[19]Roth, *supra* note 2, at 1998–99.

[20]Roth, *supra* note 2, at 2000–21.

do not present credibility issues, she states, including as an example the read out from a printer that is presented only to show that the printer was working, not for the truth of whatever the document said.[21] Similarly, she excuses tape recorders as mere "conduits" for someone else's statements, while in contrast transcription services may include errors and therefore should not necessarily be trusted as easily.[22] Finally, some machines may be only tools—such as magnifying glasses or the Thermal Cycler that copies DNA in order to aid in testing.[23]

Roth suggests that the machines above may be distinguished from machines that analyze or interpret evidence on their own, and therefore require analysis of credibility (such as a meaningful opportunity to cross examine or impeach the machine).[24] In contrast, she presents machines that do require credibility because they make independent statements that are taken for truth. Here Roth takes something of a historical view, beginning with machine statements that have, wrongly (she implies), been accepted without sufficient testing of credibility. These include photographs, which may include conscious or subconscious bias and manipulation; basic instruments such as clocks and thermometers, which may be susceptible to operators' error or manipulation but whose credibility is rarely questioned; computerized business records; and litigation software.[25]

The technology appearing in court has become more and more advanced. And as it has advanced, it has exhibited a "creeping concealedness," by offering more and more internal steps (such as completing the math a technician would once have completed and then testified to, matching fingerprints, or determining whether the amount of debris at a scene suggests a fire caused by arson).[26] As more and more steps are completed by machine, cross examination or impeachment of a human witness, or of the statement in general, becomes less and less possible. The machine cannot be cross examined, and structures are not in place to facilitate impeachment. Defendants may not even be able to question the process by which the machine comes to its conclusion, because the software may be proprietary.[27]

As machines play more and more of an independent role in gathering and analyzing evidence, Roth suggests that jurors need more

---

[21]Roth, *supra* note 2, at 2001.

[22]Roth, *supra* note 2, at 2002.

[23]Roth, *supra* note 2, at 2003.

[24]Roth, *supra* note 2, at 2005.

[25]Roth, *supra* note 2, at 2007–15.

[26]Roth, *supra* note 2, at 2016–17.

[27]Roth, *supra* note 2, at 2018.

context in order to evaluate how credible that analysis is.[28] In Section III, Roth offers a number of suggestions as to how to better restrict machine evidence and allow for more effective testing of the credibility of that evidence. She suggests front-end solutions, such as requiring any software used in litigation to maintain specific industry standards not only in design of the machines but in regular testing of the machine's accuracy.[29] She also recommends that software be open-source, so that it may be researched and tested by the public (or interested parties), and suggests that a public advisory committee might oversee key standards in proprietary software (such as what level of certainty is required before DNA software determines it has found a match or ruled one out).[30] Similarly, Roth recommends pretrial disclosure of the machine's source code so that any errors in the program can be discovered, or at least access to the " 'basic principles' underlying the machine's methods" in the case of proprietary software, and possibly pretrial access to test the machine itself.[31] Roth notes that typical authentication requirements for live testimony are often ignored in the case of machines, and even the use of *Daubert* and *Frye* reliability tests, when applied, fail to test the accuracy of machine testimony (another reason to require disclosure of information about the machine's source code).[32] Roth further recommends that any prior runs or near matches provided by the machine be disclosed to defense counsel.[33]

But, as noted above, the credibility of many machine's conclusions relies on the responsible behavior of the programmers and technicians using the machine. Roth therefore suggests not only impeachment of the machine (through the machines prior statements and/or pretrial testing of the machine as described above), but also the possibility of impeaching the machine's programmer or any person who had to input information into the machine in order to

---

[28]Roth, *supra* note 2, at 2023.

[29]Roth, *supra* note 2, at 2022–24

[30]Roth, *supra* note 2, at 2026.

[31]Roth, *supra* note 2, at 2028.

[32]Roth, *supra* note 2, at 2033–35; *compare Frye v. U.S.*, 293 F. 1013, 1014, 34 A.L.R. 145 (App. D.C. 1923) (holding that expert opinion based on a scientific technique is inadmissible unless the technique is "generally accepted" as reliable in the relevant scientific community); *with Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579, 113 S. Ct. 2786, 125 L. Ed. 2d 469, 27 U.S.P.Q.2d 1200, Prod. Liab. Rep. (CCH) P 13494, 37 Fed. R. Evid. Serv. 1, 23 Envtl. L. Rep. 20979 (1993) (reasoning that *Frye*'s general acceptance standard was overruled by the Federal Rules of Evidence and holding that the admissibility of such testimony turns on whether the underlying reasoning or methodology is scientifically valid and properly can be applied to the facts at issue).

[33]Roth, *supra* note 2, at 2029.

get a result (as required in the United Kingdom).[34] While such requirements are not made of human experts, human experts can be cross examined. The inability to cross examine machines, Roth suggests, might make testimony from machine operators more necessary.[35]

Finally, Roth turns to the Confrontation Clause, and here the crux of the article really becomes evident. She states, "[i]f the Clause is concerned with unreliable, unconfronted testimony, then credibility-dependent claims that are likely unreliable and offered against the accused at trial should pose constitutional problems, particularly if the defendant does not have the opportunity to impeach the source."[36] To fix this problem, Roth advocates for a right to meaningful impeachment of machine testimony—whether by examination of a source code, testimony by a programmer upon any updates to software, or written responses to relevant questions (such as "what threshold do you use in deciding what to call a genetic marker versus 'noise'?")[37] While not all impeachment or examination need happen in the courtroom, Roth reminds us that offering machine analysis without the ability to test that analysis may severely harm a defendant's rights to challenge the evidence presented against him.

III. Brian Sites, *Machines Ascendant: Robots and the Rules of Evidence*, 3 GEO. L. TECH. REV. 1 (2018).

In 2017, Andrea Roth suggested several remedies for scrutinizing the credibility of machine evidence. In 2018, Brian Sites surveyed judicial resolutions of arguments regarding machine evidence in order to assure us that Roth's suggestions have not been implemented, nor have any other mechanisms for judging the credibility of machine testimony. Sites argues that, "[t]rial by machine is now quite present."[38] Like Roth, Sites argues that the answer is not to exclude machine testimony, but to add scrutiny.[39] But Sites focuses on what tact the courts have taken so far, and that tact, he explains, is to reject Confrontation Clause scrutiny and use the Federal Rules of

---

[34]Roth, *supra* note 2, at 2036 (citing Stephen Mason, *Electronic Evidence, the Presumption of Reliability and Hearsay—A Proposal*, 177 CRIM. L. & JUST. WKLY. (Sept. 28, 2013), http://www.criminallawandjustice.co.uk/features/Electronic-Evidence-Presumption-Reliability-and-Hearsay---Proposal [http://perma.cc/4B9G-YLR7]).

[35]Roth, *supra* note 2, at 2037.

[36]Roth, *supra* note 2, at 2044

[37]Roth, *supra* note 2, at 2050.

[38]Sites, *supra* note 4, at 2.

[39]Sites, *supra* note 4, at 5.

Evidence to limit machine evidence instead. This strategy, he argues, has fallen short.[40]

Sites begins with a summary of the caselaw thus far, which also serves to further clarify the problems of machine testimony. The courts' general rejection of Confrontation Clause application to machine testimony begins with *United States v. Washington.*[41] In that case, a defendant was convicted of driving while intoxicated based on lab analysis of a blood sample that was taken when police pulled him over for erratic driving. The lab analysis was introduced in court via the director of the lab, without testimony from any of the actual technicians who performed the tests. The trial court held, and the Fourth Circuit confirmed, not only that the defendant had no right to cross-examine the lab technicians (he should have subpoenaed them instead), but that doing so would offer no value because the machines had made the statements, not the technicians.[42] In short, the Confrontation Clause could not apply to machines.[43] Since that time, Sites summarizes,

> A clear pattern has emerged for claims involving machine-generated data. Courts considering such claims have reached analogous conclusions for a variety of machines including those producing DNA results, breathalyzer results, urinalysis results, and machine-generated data from equipment outside the lab . . . Courts, for the most part, appear unconcerned with the rise in number of 'witnesses' immune to cross examination.[44]

Instead, defendants must rely on the Federal Rules of Evidence to test the reliability of the witnesses against them, but the Rules, he asserts, are failing in this regard. He begins, in Section II, with a basic refresher on the relevant Rules of Evidence, including authentication, authentication of a process or system, chain of custody, and the requirements for expert testimony.[45] He also addresses hearsay, but only briefly, to clarify that, like the Confrontation Clause, hearsay rules have been interpreted to apply only to statements by people.[46] Machine testimony is therefore excluded or, in rare cases, may be introduced via "a lab supervisor or other such individual testifying as a surrogate witness at trial."[47]

In Section III of his article, Sites explains where the Rules have

---

[40]Sites, *supra* note 4, at 5.

[41]*U.S. v. Washington*, 498 F.3d 225, 74 Fed. R. Evid. Serv. 332 (4th Cir. 2007); Sites, *supra* note 4, at 7.

[42]Sites, *supra* note 4, at 7–8.

[43]Sites, *supra* note 4, at 7.

[44]Sites, *supra* note 4, at 10.

[45]Sites, *supra* note 4, at 12–14.

[46]Sites, *supra* note 4, at 14.

[47]Sites, *supra* note 4, at 15.

proven inadequate to ensure reliability. He begins by outlining the series of decisions in which courts have determined that machine testimony cannot be hearsay, and that any concerns about reliability should be addressed via authentication instead.[48] Sites takes issue with this determination, arguing that the involvement of machine operators (for instance, determining test parameters) should render any statement from the machine a joint statement with the operator, and so the operator should testify.[49] Moreover, he argues (citing Roth) that machine operators may make mistakes that affect the results and that should lead to a requirement that the operators be present for cross examination and impeachment.[50]

Sites next addresses authentication. He notes that the court in *Washington*, and courts following, seem unconcerned that such damning testimony is being accepted into evidence "even where no one with first-hand knowledge testified in court that the test results were derived from the actual sample at issue."[51] Sites questions,

> How can the proponent have "produce[d] evidence sufficient to support a finding that the item is what the proponent claims it is" if the only people who know if the sample tested was the defendant's did not testify? How can the court know if the machine was operated correctly on an unadulterated sample if the operator did not testify? Similarly, how is a chain of custody demonstrated without testimony from the witness who actually analyzed the sample?[52]

And while Roth finds reason to question the lenient admission of photographs, Sites argues instead that testimony from an actual technician regarding the lab reports typically entered into evidence is far more important. While many people may have sufficient knowledge to verify the accuracy of a photograph, Sites argues that *only* the relevant machine operators or technicians know how a test was conducted or where a sample came from.[53]

Bringing together the information described above, Sites reminds us that "machine testimony is fallible,"[54] whether because it was used incorrectly, it was intentionally tampered with, the sample was altered incorrectly, or the analyst actively falsified results.[55] Each of these instances has occurred; Sites' footnotes are rife with articles from popular press or professional magazines with titles such as

---

[48]Sites, *supra* note 4, at 15–17.

[49]Sites, *supra* note 4, at 17.

[50]Sites, *supra* note 4, at 17–18.

[51]Sites, *supra* note 4, at 19.

[52]Sites, *supra* note 4, at 20 (quoting FED. R. EVID. 901(a)).

[53]Sites, *supra* note 4, at 20.

[54]Sites, *supra* note 4, at 23.

[55]Sites, *supra* note 4, at 21–22.

Court: Examine if Austin Crime Lab Botched Death Penalty Evidence; *Crime Lab Uses Wrong Chemical in 2,500 Methamphetamine Tests in Santa Clara County; How a Lab Chemist Went From 'Superwoman' To Disgraced Saboteur of More Than 20,000 Drug Cases;* and *Another Week, Another Crime Lab Scandal*.[56] Each of these scandals, from intentional tampering to accidental error, belies the assumption that machine evidence can be trusted more easily than witness testimony. Sites argues, therefore, that "a critical right for defendants is the ability to challenge, through cross-examination of the individual(s) who ran the test, whether the defendant's actual sample was tested and tested properly."[57]

IV. CONCLUSION

In recent years we have learned that we have had too much faith in criminal forensics, as evidenced by the scandals listed above as well as many other discoveries of inaccuracy.[58] Yet, as Sites points out, the increasing use of advanced technology is leading the courts to accept this evidence with *less* oversight, and *fewer* opportunities for defendants to challenge the credibility of the accusations against them. As Roth argued that more oversight is needed, Sites adds his voice not only to say that oversight is needed, but that the trend is in the opposite direction. This choice to "exemp[t] nearly all machine statements [from hearsay provision in the Federal Rules of Evidence] is disingenuous and fails to reflect the reality . . . These or other analytical changes—such as reinterpreting the Confrontation Clause—are increasingly important as machine accusers aided by human operators rise in prevalence."[59]

The threat posed by this trend is a complete loss of the right to face one's accusers and challenge the prosecution's evidence. As Sites argues, "[w]hen the prosecution fills its evidence list with machine accusers and their human supervisors . . . who have no first-hand knowledge of the analysis allegedly conducted—how will

---

[56]Sites, *supra* note 4, at 21–22, nn.130–33.

[57]Sites, *supra* note 4, at 24.

[58]Radley Balko, *We Need to Fix Forensics. But How?*, WASH. POST (June 20, 2019), https://www.washingtonpost.com/opinions/2019/06/20/we-need-fix-forensics-how/; Radley Balko, *Two FBI Officials Say the State of Forensics Is Fine. Here's Why They're Wrong.* WASH. POST (June 6, 2018), https://www.washingtonpost.com/news/the-watch/wp/2018/06/06/two-fbi-officials-say-the-state-of-forensics-is-fine-heres-why-theyre-wrong/; Ryan Gabrielson, *The FBI Says Its Photo Analysis Is Scientific Evidence. Scientists Disagree*, PROPUBLICA (Jan. 17, 2019), https://www.propublica.org/article/with-photo-analysis-fbi-lab-continues-shaky-forensic-science-practices; Leora Smith, *How a Dubious Forensic Science Spread Like a Virus*, PROPUBLICA (Dec. 13, 2018), https://features.propublica.org/blood-spatter-analysis/herbert-macdonell-forensic-evidence-judges-and-courts/.

[59]Sites, *supra* note 4, at 25.

the defendant prove the machine erred or was influenced to err?"[60] At stake is only the fundamental premise of the adversarial system. It would be better if we did not simply ignore the issue.

---

[60]Sites, *supra* note 4, at 27.