April 11, 2024

Dear LLS faculty workshop participants:

This short piece summarizes some proposals I've made that will be considered by the Advisory Committee on the Rules of Evidence on Friday, April 19, as well as some other proposals for concrete statutory changes to rules of evidence and criminal procedure I've been mulling over for a while. All feedback welcome, and I'll be sure to incorporate it into the discussion before the Advisory Committee as well.

Best,
Andrea

Andrea Roth
Barry Tarlow Chancellor's Chair in Criminal Justice and Professor of Law
UC Berkeley School of Law
aroth@law.berkeley.edu

PROPOSED NEW EVIDENCE AND PROCEDURE RULES FOR MACHINE CONVEYANCES OF INFORMATION

Andrea Roth[1]

This short essay briefly explains how machine-generated conveyances of information are now ubiquitous as proof, how they raise reliability concerns similar (though not identical) to human assertions, and how existing rules of evidence and procedure fail to regulate these machine conveyances in a way that would offer safeguards similar to those governing human assertions. It then offers several concrete proposals, some of which will be discussed by the Advisory Committee on the Rules of Evidence at its April 2024 meeting, to better ensure that machine-generated conveyances of information do not threaten the accuracy of verdicts.

I. The Significance of Machine Conveyances of Information in Modern Trials

Some proof offered in trials is generated electronically, by a machine or algorithm, and creates and conveys output in the form of information, in a way that would be testimony or hearsay (a statement offered to prove the truth of the matter asserted) if uttered by a human. In previous work, I have called this sort of proof "machine testimony," distinguishing it from electronically *stored* information (like an email) or machine tools (like a robot that performs a physical task like filling test tubes). These other non-testimonial forms of machine-facilitated proof might raise important issues of authentication or chain of custody, but they do not raise the same sort of concerns that machine conveyances of information do.[2] Deep fake videos, for example, are a serious problem for evidence law, but mostly one of authentication (how do we know this *thing* is what the proponent claims it to be – e.g., the president of Ukraine telling his people to surrender?). In contrast, a likelihood ratio offered by DNA software is what it claims to be (a likelihood ratio offered

Machine conveyances of information are now routine fixtures in both pretrial criminal investigations and civil and criminal trials.[3] Of course, the readings of basic scientific instruments have been admitted in trials for centuries. But in the computer age, the amount of machine-generated proof has significantly grown. With respect to pretrial investigations, for example, the results of algorithms involving facial recognition, voice recognition, and DNA phenotyping software are used to identify suspects for further surveillance and

[1] Professor of Law and Barry Tarlow Chancellor's Chair in Criminal Justice, UC Berkeley School of Law.

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

[3] Algorithms have obviously played a large role in dangerousness prediction in both the pretrial and post-conviction context as well, including with the blessing of many progressives in the name of eliminating money bail. Others have ably explained the bias and accuracy problems with using algorithms for dangerousness prediction, and their limited utility other than perhaps as a mirror to expose the problems with risk assessment as a "neutral" principle in general. *See, e.g.*, Colin Doyle, *The Feature Is the Bug*, INQUEST, Aug. 9, 2021. I therefore do not focus on legislative or rule-based solutions to risk assessment algorithms in this project.

potential arrest.[4] With respect to proof of guilt/innocence (or liability/non-liability) at trial, parties have offered Google Earth location estimates, driving time estimates, likelihood ratios for potential DNA mixture contributors generated by probabilistic genotyping software, "supercharged" video footage enhanced with images predicted by machine learning software;[5] blood-alcohol concentration estimates from software-driven machines, conclusions of machine-learning classifier algorithms as to which of three people likely wrote a tweet confessing to a homicide, results of automated forensic software for voice recognition, Fitbit data offered to determine whether someone was sleeping at a particular time, time-stamp data on photographs, license plate readers, address logs purporting to list IP addresses of users who have visited a particular website, and Event Data Record information.

On the horizon (that is, technologies that already exist though they have not yet been introduced at trial) include algorithms that assess verbal eyewitness confidence statements through natural language processing;[6] DeepPlate, a deep-learning-based algorithm for "decyphering" license plates from blurry photographs;[7] AI-based medical diagnoses and cause-of-death determinations;[8] and Large Language Models (LLMs) that identify attempted fraud.[9] Indeed, a recent report by Europol suggested that police agencies create their own custom LLMs for purposes of detecting particular types of criminal activity.[10] ChatGPT itself reports that it "[c]an analyze text, such as a transcript of a confession, and provide insights into truthfulness based on linguistic cues, writing style, and content," although it warns that its conclusions "should not be solely relied upon for definitive authorship attribution." Researchers have also developed an Automated Neural Nursing Assistant[11] that determines,

---

[4] *See, e.g.*, Kashmir Hill, Eight Months Pregnant and Arrested After False Facial Recognition Match, N.Y. Times, Aug. 6, 2023 (noting false arrest of Porcha Woodruff based on facial recognition technology); Andrew Pollack, Building a Face, and a Case, on DNA, N.Y Times., Feb. 23, 2015 (noting use of DNA phenotyping to create a composite facial sketch for use in identifying a suspect in an unsolved South Carolina murder); Dep't of Homeland Security, Snapshot: Voice Forensics Can Help the Coast Guard Catch Hoax Callers, Sept. 26, 2017, available at https://www.dhs.gov/science-and-technology/news/2017/09/26/snapshot-voice-forensics-can-help-coast-guard-catch-hoax (noting use of voice forensics in investigation to determine source of false distress call).

[5] *See* State v. Joshua Everybodytalksabout, No. 21-1-04852-2, Findings of Fact and Conclusions of Law Re: Frye Hearing on Admissibility of Videos Enahnced by Artificial Intelligence (King C'ty, Wash. Mar. 29, 2024).

[6] *See* Rachel Leigh Greenspan et al., Penn Law Public Law and Legal Theory Research Paper Series Research Paper No. 24-06, PSYCHOL. SCI. (forthcoming 2024), available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4720985.

[7] *See Introducing DeepPlate, Amped's Investigative Tool for AI-Powered License Plate Reading,* https://blog.ampedsoftware.com/2024/02/28/introducing-deepplate-ampeds-investigative-tool-for-ai-powered-license-plate-reading (although offering the caveat that the tool "is not currently reliable for legal evidence").

[8] *See, e.g.*, Mugahed A. Al-Antari, Artificial Intelligence for Medical Diagnostics—Existing and Future AI Technology, 13 DIAGNOSTICS 688 (2023), available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9955430/#:~:text=AI%20algorithms%20can%20analyze%20medical,diseases%20more%20accurately%20and%20quickly.

[9] Scanlon et al., ChatGPT for digital forensic investigation: The good, the bad, and the unknown, Forensic Science International: Digital Investigation 46 (2023) 301609.

[10] Europol (2023), ChatGPT - The impact of Large Language Models on Law Enforcement, a Tech Watch Flash Report from the Europol Innovation Lab, Publications Office of the European Union, Luxembourg (""Law enforcement agencies may want to explore possibilities of customised LLMs trained on their own, specialised data, to leverage this type of technology for more tailored and specific use.").

[11] *See* Automated Neural Nursing Assistant (ANNA): An Over-The-Phone System for Cognitive Monitoring, Jacob Solinksy et al., Proceedings of the Annual Conference of the International Speech Communication Association,

based on a series of conversations, whether a defendant is mentally deteriorating, raising the potential for AI to soon opine on a defendant's competence to stand trial, malingering, or the like.

II. Reliability Concerns Raised by Machine Conveyances

While the rise in machine conveyances of information holds much promise for the legal system in terms of enhanced accuracy in detection of wrongdoing and identification of perpetrators, it also raises significant potential reliability concerns. In previous work, I and others have catalogued numerous instances of false conveyances of information by machines, stemming from deliberate deceptions, coding errors, questionable analytical assumptions, skewed datasets, electronic malfunctions, and the like.[12] These "black box dangers" are analogous to the so-called hearsay dangers that might lead a factfinder to draw a false inference from a human witness's assertion – insincerity, ambiguity/inarticulateness, misperception, and memory loss.[13]

Three recent examples illustrate these various concerns. First, facial recognition software has resulted in numerous false arrests, including three false arrests in Detroit alone in the last three years, all of African Americans.[14] As detailed in a recent white paper issued by Georgetown Law's Center on Privacy and Technology, these inaccuracies appear to be the result of incomplete and biased training data that do not sufficiently represent people of color, as well as photo quality problems and confirmation bias from the facial match tainting subsequent human identifications.[15]

Another example relates to AI-enhanced imaging. Two weeks ago, a Washington State man accused of murder, Joshua Puloka, tried to introduce a cell phone video of the deadly fight that led to the charges. According to Puloka, the video shows he tried to deescalate the violent situation and shot in self-defense, accidentally killing his alleged assailant and two bystanders. The version of the video offered by Puloka was enhanced, or "supercharged," by a machine learning program, Topaz Labs' "Video AI 4." The prosecution moved to exclude the video on grounds that it was "inaccurate, misleading, and unreliable." In particular, the prosecution objected that the program "utilized an opaque, proprietary process which no witness can explain," and its expert further argued that the enhanced video changed the colors of participants' clothes and shoes, showed a sharp object protruding from a man's leg that in the original video was only a shadow, and the like. Puloka's lawyers insisted, in response, that the enhanced version was a simple machine learning classifier, just like

---

Aug. 2023, available at https://experts.umn.edu/en/publications/automated-neural-nursing-assistant-anna-an-over-the-phone-system-.

[12] Roth, *Machine Testimony*, at 1978, 1989–90. *See also* Brandon Garrett & Cynthia Rudin, *The Right to a Glass Box: Rethinking the Use of Artificial Intelligence in Criminal Justice*, 109 CORNELL L. REV. -- (forthcoming 2023).

[13] See Edmund M. Morgan, Hearsay Dangers and the Application of the Hearsay Concept, 62 HARV. L. REV. 177, 188 (1948).

[14] *See* After Third Wrongful Arrest, ACLU Slams Detroit Police Department for Continuing to Use Faulty Facial Recognition Technology, ACLU, available at https://www.aclu.org/press-releases/after-third-wrongful-arrest-aclu-slams-detroit-police-department-for-continuing-to-use-faulty-facial-recognition-technology.

[15] *See generally* Georgetown Law, Center on Privacy and Technology, A FORENSIC WITHOUT THE SCIENCE: FACE RECOGNITION IN U.S. CRIMINAL INVESTIGATIONS, DEC. 6, 2022, available at https://www.law.georgetown.edu/privacy-technology-center/publications/a-forensic-without-the-science-face-recognition-in-u-s-criminal-investigations/.

numerous widely used AI tools today, and was a "faithful depiction of the original." The trial judge ultimately excluded the video under the so-called "*Frye* rule" for novel scientific evidence, requiring such evidence to have gained "general acceptance" in the relevant scientific community. The judge reasoned that the program used novel "opaque methods" and a "non-peer-reviewable-process" to "represent what the AI model 'thinks' should be shown" and had not yet been accepted as accurate.[16]

Finally, consider the opposing results of two competing DNA software programs in a murder case a few years ago in upstate New York involving the strangulation of a 12-year-old boy. A couple of ex-boyfriends of the victim's mother were suspects. One had a history of domestic violence but an apparent alibi. The other was a soccer coach at a local college named Nick Hillary. But there was no apparent motive for either suspect, nor was there any physical evidence linking them or anyone else to the crime.[17] There was, however, a small amount of DNA from a minor contributor on a fingernail scraping from the boy, which was too complex a mixture for human analysts. The district attorney first sent the mixture to an American company with a probabilistic genotyping program called TrueAllele. Based on TrueAllele's deconvolution of the mixture, the company's CEO, Mark Perlin, reported there was "no statistical support" for concluding that Mr. Hillary was the minor contributor. Indeed Mr. Perlin would later insist the evidence suggested Mr. Hillary was *excluded* as a contributor, based on the likelihood ratios he calculated.[18] Eventually the DA sent the same DNA mixture information to a second company, a New Zealand competitor to TrueAllele, called STRMix. In contrast to TrueAllele, STRMix reported a highly damning likelihood ratio of over 300,000 with respect to Mr. Hillary (stating that the mixture was 300,000 more likely if Hillary was a contributor than if he was not), based on which he was arrested and prosecuted for the murder. Eventually, the trial judge ruled the STRMix results inadmissible because the software had not been sufficiently internally validated by the local crime laboratory, and Mr. Hillary was acquitted. After the verdict, the creator of STRMix issued a memorandum explaining the differing assumptions underlying the STRMix and TrueAllele results and explaining why the STRMix conclusion was the correct one.[19] The creator of TrueAllele did the same, arguing that STRMix would also have excluded Hillary had it chosen a different analytical threshold for which data to consider.[20]

The point of this tragic case for purposes of this essay is not that Mr. Hillary is necessarily guilty or not guilty, but that two purportedly reliable probabilistic genotyping software programs came to two vastly different conclusions, at least as reported, about the same DNA mixture and the same criminal suspect, in a high stakes case. One or both of the programs must have created false or misleading information, in a way that might have gone unnoticed had the other program's result not been known.

---

[16] *See* State v. Joshua Everybodytalksabout, No. 21-1-04852-2, Findings of Fact and Conclusions of Law Re: Frye Hearing on Admissibility of Videos Enahnced by Artificial Intelligence (King C'ty, Wash. Mar. 29, 2024).

[17] The case is the subject of a recent HBO documentary miniseries, Who Killed Garrett Phillips?. *See* https://www.hbo.com/who-killed-garrett-phillips/season-1.

[18] *See, e.g.*, Mark Perlin, *STRMix v. Buckleton: Misinterpretation of DNA Evidence in People v. Oral Hillary*, July 29, 2016, available at https://www.cybgen.com/information/newsroom/2016/aug/files/STRvJSB.pdf.

[19] *See* John Buckleton, *People v. Hillary*, STRMix, Dec. 2017, available at https://johnbuckleton.wordpress.com/wp-content/uploads/2017/12/people-v-hillary-ii.pdf.

[20] Perlin (2016) at 2 (arguing that the 30 versus 50 RFU cutoff made the difference).

III. The Underregulation of Machine Conveyances by Existing Rules of Evidence and Procedure

The mere fact that a type of evidence raises reliability concerns does not mean it needs to be regulated by the rules of evidence. As one of my mentors, the late Eleanor Swift, always reminded me, the purpose of evidence rules is not to sanitize the courtroom of all unreliable evidence, but to ensure that the jury has the contextual information it needs to accurately assess the probative value of the evidence.[21] Hence, we allow witnesses with six perjury convictions to testify, even if they are compulsive liars; the prior conviction information and the jurors' own life experiences are likely sufficient to allow them to assess the evidence appropriately without drawing the wrong inference.

When it comes to *human* conveyances of information, the Anglo-American rules of evidence generally require that they be offered on the witness stand, under oath and subject to cross-examination, so that the opponent can discover and expose potential infirmities and incomplete aspects of a witness's claims. These requirements are enforced through the rule against hearsay, which generally excludes out of court assertions of human declarants when offered for their truth unless they come within an exception to the rule (such as for business records or "excited utterances"), as well as the hearsay rule's constitutional cousin, the Confrontation Clause of the Sixth Amendment (guaranteeing an accused the right "to be confronted with the witnesses against him). Although courts and commentators love to remind litigants and the public that there is no general constitutional right to "discovery," the rights to physical confrontation and cross-examination enforced through the hearsay rule and Confrontation Clause might be most naturally thought of as rules of discovery. After all, one reason confrontation is critical is not simply to offer a means of exposing mistakes and lies (through impeachment) or discouraging mistakes and lies (through the ennobling power of the oath), but of discovering them as well – the supposedly telltale beads of sweat on the brow of a liar; the exploration of broader explanatory circumstances of an ambiguous or isolated statement taken out of context; the eliciting of a recantation upon backing a witness into a corner by confronting them with inconsistencies or tugging at their guilty conscience. Put simply, we definitely constitutionalize certain forms of discovery; the question is why cross-examination and physical confrontation should be privileged above other forms.

Beyond just the infamous hearsay rule and Confrontation Clause, plenty of other statutory and common-law rules of evidence and procedure contain safeguards to avoid wrongful convictions from unreliable witness testimony, such as: rules of basic competence (e.g. Federal Rules of Evidence 601 and 602, requiring that witnesses take and understand an oath to tell the truth, and speak only based on personal knowledge), court rules of discovery (e.g. Federal Rule of Criminal Procedure 16 and Civil Rule 26, requiring information about expert witnesses to be disclosed before trial, to aid in scrutinizing expert claims); the Jencks Act and Federal Rule of Criminal Procedure 26.2 (requiring parties in criminal cases to disclose all prior recorded statements of their witnesses on the subject matter of their testimony); rules of impeachment (such as Federal Rules of Evidence 608, 609, 613, and 801(d), allowing impeachment of both testifying witnesses and hearsay declarants with evidence of dishonesty and inconsistencies, and common-law doctrines

[21] *See generally* Eleanor Swift, *A Foundation Fact Approach to Hearsay*, 75 CALIF. L. REV. 1339 (1987) (arguing for an approach to hearsay that focuses on giving factfinders sufficient context about a statement's meaning, rather than on excluding unreliable assertions).

allowing impeachment with evidence of bias and incapacity); corroboration (such as rules disallowing a conviction based on uncorroborated confessions, accomplice testimony,[22] or a sole witness alleging perjury or treason); jury instructions; and front-end safeguards related to creation and preservation of evidence.[23]

When it comes to machine conveyances of information, however, the jury often has little if any context with which to judge their probative value accurately. For example, imagine a criminal defendant in federal court charged with a crime, where the primary evidence of guilt is the following conclusion of a DNA software program: "There are 3 contributors to the DNA mixture on the gun, and based on the DNA typing results obtained, it is at least 49 million times more likely if the observed profile from the swabs of the textured areas of GUN-001 originated from [Defendant] and two unrelated, unknown contributors than if the data originated from three unrelated, unknown individuals."

If this statement were offered into evidence without an accompanying human expert, it would be subject only to requirements of relevance (FRE 402, requiring that evidence be probative of a material fact to be admissible) and authenticity (FRE 901, 902, requiring that the proponent offer proof sufficient to show that the evidence is what the proponent claims it to be). But relevance and authenticity are easily met requirements. Evidence is relevant so long as it "has any tendency to make a fact" that matters to the case slightly "more or less probable than it would be without the evidence." Even evidence that has a high error rate is typically at least relevant if it has anything to do with something that matters to the case, such as identification. In terms of authentication of machine-generated the showing is also minimal and can now even be shown without a live witness.[24] To be sure, authentication for machine-generated proof oddly requires that the proponent show not only that the machine output is what the proponent says it is but also that the process or system produces an "accurate result."[25] The original language of the rule for authenticating machine-generated proof drafted in 1968 provided, much like other traditional authentication rules, that a proponent prove that a system result "fairly represents or reproduces the facts which the process or system purports to represent, or reproduce." But Judge Jack Weinstein suggested adding the word "accurate" to the language as a means of dealing with output of IBM counting machines, giving proponents maximum flexibility. In short, while Rules 901 and 902 seem to require proof of an algorithm's accuracy, there is no meat on the bones of this rule and it requires only proof that a reasonable juror might accept as sufficient.

On the other hand, if the statement from this DNA software were offered into evidence in federal court with a human expert relying on it to render an opinion, then the expert's opinion would at least be subject to the reliability requirements of FRE 702 and the so-called "*Daubert* standard," requiring the proponent to show by a preponderance of the evidence that

---

[22] *See, e.g.,* N.Y. Crim. Proc. Law § 60.22 (McKinney 2016) (prohibiting a criminal conviction "upon the testimony of an accomplice unsupported by corroborative evidence").

[23] *See, e.g.*, State v. Henderson, 27 A.3d 872, 878 (N.J. 2011) (establishing protocols for eyewitness identification procedures). Other examples of rules of production governing human assertions include state laws requiring that confessions be videotaped.

[24] *See* FRE 902(13) (allowing self-authentication, by certificate, of a "record generated by an electronic process or system that produces an accurate result, as shown by a certification of a qualified person that complies with the certification requirements of Rule 902(11) or (12)."

[25] *See* FRE 901(9), 902(13).

the software program that produced the expert's testimony is a "reliable . . . method[]."[26] *Daubert* hearings on machine-generated proof generally boil down to examination of existing validation studies and competing affidavits or testimony from experts as to the potential problems with the software. Validation studies are important, to be sure, but have three limitations as a means of ensuring that factfinders have enough context to judge the probative evalue of algorithmic output. First, validation studies typically speak to the potential for false positives (because they are studies conducted with a known ground truth), but not so much to the reliability of the program's reported "scores" (such as likelihood ratios).[27] Second, validation studies are small in number and typically involve run-of-the-mill samples rather a large set of studies and samples showing the limits of the method over a large multi-vectored "factor space."[28] Third, software output is typically from a proprietary process where the creator claims the source code as a trade secret and validation studies are conducted by the proprietor themselves or a financially dependent entity. These algorithms are not subject to the sort of stress testing suggested by the Institute of Electrical and Electronics Engineers (IEEE) for high stakes algorithms (like the ones used in banking or powering the elevators in the law school).

Put more simply, in terms of scrutinizing the claims of human experts, would we ever allow an expert to testify via affidavit rather than live testimony, so long as their affidavit was accompanied by a small pile of validation studies showing that the expert's method is minimally reliable? No. We would still require all the other testimonial safeguards to be provided to the opponent (discovery, impeachment, cross-examination). And in civil cases, we might also subject the expert to a pretrial deposition and interrogatories. In criminal trials, the opponent's first crack at the expert is at trial, unless the expert is voluntarily willing to speak with the opposing party beforehand. But with respect to algorithms, proprietors routinely deny access to their programs for use by academic researchers. My colleagues and I at Berkeley have tried in vain to obtain basic research licenses for the two main DNA software programs, STRMix and TrueAllele, to conduct independent audits of our own, to no avail.

---

[26] FRE 702(c). The Supreme Court in a series of cases in the 1990s dubbed the "*Daubert* trilogy" interpreted FRE 702 to require proof of an expert method's reliability, both foundationally and as applied. [*Daubert v. Merrill Dow Pharm. Co.* (1993); *General Electric v. Joiner* (1997); *Kumho Tire v. Carmichael* (1999)]. Rule 702's language now incorporates the holdings of these cases. Most states now follow the *Daubert* standard, although a minority of states, including California, continue to follow the *Frye* "general acceptance" standard that applies only to novel scientific evidence and that asks judges not to determine scientific validity directly themselves, but rather determine whether the scientific community itself believes the method to be reliable. *But see* Sargon Enterprises, Inc. v. University of Southern California, 55 Cal. 4th 747 (2012) (requiring judges to exercise a "gatekeeping" function and exclude expert methods that are "clearly . . . unreliable" because based on speculation or lack of "intellectual rigor").

[27] *See* Christopher D. Steele & David J. Balding, Statistical Evaluation of Forensic DNA Profile Evidence, 1 Ann. Rev. Stat. & Its Application 361, 380 (2014) (arguing that "validation" against a ground truth "is infeasible for software aimed at computing a[] [likelihood ratio] because it has no underlying true value (no equivalent to a true concentration exists). The [likelihood ratio] expresses our uncertainty about an unknown event and depends on modeling assumptions that cannot be precisely verified in the context of noisy [crime scene profile] data").

[28] See NIST, DNA Mixture Interpretation: A NIST Scientific Foundation Review (2021), available at https://www.nist.gov/news-events/news/2021/06/nist-publishes-review-dna-mixture-interpretation-methods) (emphasizing the importance, in determining software accuracy in a given case, of validation studies covering the particular "factor space" that a case falls into (such as, in the DNA mixture context, the quantity of DNA, number of contributors, peak height differential, extent of relatedness and thus allele-sharing among contributors, etc.).

Of course, when a human expert himself takes the stand, he is presumably available to answer questions about the program and its inputs and assumptions. But this safeguard, too, has limitations. First, some machine outputs are offered without a human expert attached, and so long as the output is authenticated and is relevant, nothing stops a proponent from doing so.[29] Second, some machine processes are becoming so complex so rapidly that a programmer cannot meaningfully explain, except in very broad strokes, why the program is doing what it is doing. Third, we would never allow such testimony to substitute for meaningful scrutiny of the human expert who actually did the analysis and made the claim.[30]

Meanwhile, machine testimony is not subject to any of the other safeguards we have for human testimony, in terms of competence, discovery rights (other than where the program is used by a human expert, whose own testimony is subject to discovery rules), access to Jencks material, or corroboration requirements. While breath-alcohol machines are subject to meaningful front-end safeguards under many state laws and the Code of Federal Regulations (listing approved machines and required protocols for use), they are an outlier (and likely the result of the unusual political capital of defendants charged with DUI compared to other criminal defendants). And while a party might be able to offer some impeachment evidence against a machine or an algorithm's creator, such as to show bias or a prior inconsistency, current rules might reasonably be construed to disallow impeachment of a machine conveyance with evidence of a prior false allegation.[31]

In sum, the primary gaps in the rules of evidence with respect to machine-generated proof are:

1. Little to no scrutiny of the reliability of machine-generated proof when it is not accompanied by an expert witness's testimony (and thus not subject to FRE 702);
2. Limited scrutiny of the reliability of machine-generated proof even when it is the method underlying an expert witness's testimony (and thus subject to FRE 702), because the primary evidence relied on is often validation studies limited in scope, number, and independence from the software proprietor;
3. Limits rights of impeachment and discovery (including potential pretrial access) analogous to such rights with respect to lay and expert human assertions;
4. No rules of competence or corroboration or front-end safeguards required for admissibility, outside the context of breath-alcohol machines.

IV. Suggested New Rules of Evidence and Procedure for Machine Conveyances of Information

[29] *See, e.g.*, People v. Lopez, 286 P.3d 469, 472 (Cal. 2012) (portion of spectrometer readings offered to prove blood-alcohol concentration, without additional sworn testimony or certification by human expert, not subject to the Confrontation Clause because not the statement of a human expert, over dissent by Justice Liu).

[30] *See* Bullcoming v. New Mexico, 564 U.S. – (2011) (holding that testimony of a surrogate witness who did not perform the blood-alcohol analysis or write the hearsay report offered into evidence was not a sufficient substitute under the Confrontation Clause for the live testimony of the analyst who actually authored the report).

[31] *See* Fed. R. Evid. 608(b) (excluding extrinsic evidence of prior instances of untruthfulness to impeach a witness, though allowing such evidence on cross-examination of a live human witness); Fed. R. Evid. 806 (allowing impeachment of hearsay declarants, including by inconsistency even if the declarant was not confronted with the inconsistency, but offering no way to get around 608(b)); Nevada v. Jackson, 569 U.S. 505, 511 (2013) (upholding denial, on AEDPA grounds, of criminal defendant's request to impeach hearsay declarant with prior false allegation).

A. Federal Rules of Evidence

The first change to the Federal Rules of Evidence I am suggesting to the Advisory Committee is to Rule 702, along the lines of the following:

702. Testimony by Expert Witnesses.
(1) A witness who is qualified as an expert by knowledge, skill, experience, training, or education may testify in the form of an opinion or otherwise if the proponent demonstrates to the court that it is more likely than not that:

(a) the expert's scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue;

(b) the testimony is based on sufficient facts or data;

(c) the testimony is the product of reliable principles and methods; and

(d) the expert has reliably applied the principles and methods to the facts of the case.

(2) Where the output of a process or system would be subject to part (1) if testified to by a human witness, the proponent must demonstrate to the court that it is more likely than not that:

(a) The output will help the trier of fact to understand the evidence or to determine a fact in issue;

(b) The output is based on sufficient and pertinent inputs and data, and the opponent has reasonable access to those inputs and data;

(c) The output is the product of reliable principles and methods; and

(d) The output reflects a reliable application of the principles and methods to the facts of the case, based on the process or system's demonstrated reliability under circumstances or conditions substantially similar to those in the case.

(3) The output of basic scientific instruments and tools are not subject to the requirements of this rule.[32]

Alternatively, the Rules Committee could leave Rule 702 as is (because it was just amended in 2023 to explicitly require the preponderance standard, and apparently the Committee does not like to tinker with a rule two years in a row) and add a new rule along the lines of the following (I thank Dan Capra for this suggestion):

707. Machine-Generated Evidence.
Where the output of a process or system would be subject to Rule 702 if testified to by a human witness, the court must find by a preponderance of the evidence that:

(a) the output satisfies the requirements of Rule 702;

the output is the product of a process or system with demonstrated reliability under circumstances or conditions substantially similar to those in the case.

The output of basic scientific instruments and tools are not subject to the requirements of this rule.

[32] Basic scientific instruments have been regulated by courts through common-law rules since the 19th century; courts treat them similarly to dog alert evidence and require a showing of reliability and working operating condition. The rules already have similar exceptions that seem to work, such as hearsay exceptions for "learned treatises" and such.

I have also submitted a proposal to the Advisory Committee to amend Rule 806 as follows:

Rule 806. Attacking and Supporting the Declarant

(1) When a hearsay statement — or a statement described in Rule 801(d)(2)(C), (D), or (E) — has been admitted in evidence, the declarant's credibility may be attacked, and then supported, by any evidence that would be admissible for those purposes if the declarant had testified as a witness. The court may admit evidence of the declarant's inconsistent statement or conduct, regardless of when it occurred or whether the declarant had an opportunity to explain or deny it. If the party against whom the statement was admitted calls the declarant as a witness, the party may examine the declarant on the statement as if on cross-examination.

(2) When output of a process or system has been admitted in evidence, and would be a hearsay statement if uttered by a human declarant, the output's accuracy may be attacked, and then supported, by any evidence that would be admissible for those purposes if the output had been uttered by a human declarant. The court may admit evidence of the process or system's inconsistent output, where such output would be admissible under 806 if the output were hearsay. The court may also admit prior false output where probative of the admitted output's accuracy and where the opponent offers evidence sufficient to support a finding of the prior false output. [*Where the court admits such prior false output, the court shall inform the factfinder of the content of the prior false output, but shall not allow further evidence from either party on the matter.]*

Dan Capra, the reporter for the Advisory Committee, agrees with the thrust of my proposal but was concerned that applying Rule 806 wholesale to machines might be confusing or create unintended consequences. Indeed, Professor Capra himself in 2021 submitted a suggested amendment to Rule 806 to allow hearsay declarants to be impeached with evidence of prior false allegations (which is currently not allowed, because of 608(b)'s prohibition on extrinsic evidence of prior instances of dishonesty other than convictions, a rule concerned with avoiding mini-trials over disputed events long ago). He notes that such an amendment creates the awkward result that hearsay declarants are subject to more powerful impeachment (and potential mini-trials) even than live witnesses. In my view, the proponent's reliance on hearsay is what would justify this asymmetry; the less context the jury has because of the lack of live testimony, the more flexible the rules should be with respect to impeachment. Moreover, Professor Capra came up with an elegant solution to his own identified problem, which would be to simply note for the jury the prior false allegation, rather than allow a mini-trial. I like this suggestion, which I incorporate in the above bracketed italicized language in green (but which I haven't yet submitted to the Committee).

Finally, I have submitted a proposal to amend Rule 901(b)(9), which currently governs admissibility of output of a process or system, requiring proof that it produces an accurate result. As I noted earlier, this requirement is really a minimal add-on to authenticity and was not meant, in 1968, to be a robust reliability requirement for complex algorithms. Still, Rule 901 might be a convenient and natural place to add whatever additional conditions of admissibility we might think are important for machine conveyances.

Of course, the conditions included below could also simply be required by statute, completely separate from the rules of evidence. For example, as Brandon Garrett and Cynthia Rudin have recently written in a piece calling for "Glass Box AI":

> We propose that legislation require glass box or interpretable AI be mandatory for most uses by law enforcement agencies in criminal investigations. So long as the use of AI could result in material or information used to investigate and potentially convict a person, it should be fully interpretable. Further, all law enforcement systems should be validated, based on adequate data. Validation and interpretability should be required by statute.[33]

In addition, Congressman Mark Takano (D-CA) has tried multiple times to get through a bill, the Justice in Forensic Algorithms Act, that would require greater transparency and discovery with respect to algorithms used in criminal justice applications. But this bill appears to be dead, with little chance of revival. Let's face it – getting Congress to focus on what might appear to be an unprecedented progressive expansion of discovery rights with respect to AI might be highly controversial, even for those who might sympathize with the cause. In contrast, if these sorts of requirements were seen as a logical and modest extension of rules that already exist, and if the wordsmithing and internal debates occurred in the Advisory Committee to rules of evidence and procedure rather than in Congress, I think they would be more likely to be adopted and probably higher quality.

Some of the requirements below would be duplicative of the amendments to Rule 702 above, so if that were implemented, this list could be tweaked. The point is to not have algorithms fall through the cracks where they are not accompanied by a human interolocutor, and even when they are, to require these additional findings that are uniquely applicable to algorithms and not humans.

901(b)(9). Evidence About a Process or System. Evidence describing a process or system and showing that it produces a~~n accurate~~ reliable result, including, with the exception of basic scientific instruments, a showing of all of the following:

(A) that the opponent had adequate pretrial access to the process or system;

(B) in a criminal case, the proponent has disclosed all previous output of the process or system that, if the process or system were a human witness, would be disclosable under 18 U.S.C. §3500;[34]

(C) that the process or system has been shown through testing by a financially and otherwise independent entity to produce an accurate result under conditions substantially similar to the instant case;

[33] Garrett & Rudin (2023), *supra* note *, at *.

[34] This is the Jencks Act, requiring disclosure by the end of the direct examination of all prior substantially verbatim recorded statements of testifying witnesses on the same subject matter as the testimony. I think it's obvious that the Jencks Act should also apply to human hearsay declarants, which, bizarrely, it does not. I've lost that appellate issue in the D.C. Court of Appeals, and it's somewhat beyond the scope of this essay, but I think it's a no-brainer.

(D) that the process or system, or a license to use it, is accessible to independent research bodies, including the National Institute of Standards and Technology and accredited educational institutions, for purposes of conducting audits of the process or system;

(E) that the process or system is either open source or the proprietor has given the National Institute of Standards and Technology access to its source code for the limited purpose of conducting audits consistent with the proprietor's intellectual property rights under any applicable laws;

(F) that, in a criminal case, the proponent has not invoked a trade secrets privilege to block access or disclosure to the process or system, its source code, or the data on which it relies.

This list could be added to. For example, I did not include a requirement that the algorithm be subject to the IEEE's requirements for stress testing of high-stakes algorithms, nor that the proprietor actually turn over the source code as a condition of admissibility, nor that the algorithm be open source or "interpretable" (as Garrett and Rubin suggest). I think some of those are unrealistic or potentially even bad policy with unintended consequences (e.g. an open source requirement, which might exclude Google Earth estimates even if they were, say, critical to the jury's fair assessment of a defendant's alibi), and others are perhaps too vague or complex for a rule of evidence (e.g. whether the algorithm is sufficiently "interpretable"). I also did not require unfettered access to data sets on which classifier algorithms are trained; I think this is a critical issue but there are difficult privacy issues to work through. Your suggestions are welcome on this front.

B. *Rules of Criminal Procedure*

The above proposed amendments to the rules of evidence would accomplish much of what I think would be necessary in terms of facilitating adversarial testing of algorithms. They could also be accomplished as rules of procedure, such as amending Criminal Rule 26.2 (the Jencks Act, as applied to pretrial hearings) to require Jencks material of machines), or amending Criminal Rule 16 to require certain disclosures (currently required above in 901(b)(9) as conditions of admissibility of expert systems. Perhaps there are strategic or logistical reasons that including these requirements in the rules of procedure would be better or easier than including them in the rules of evidence.

In another forthcoming essay, I explore the possibility of a two-machine corroboration requirement, inspired by the Oral Hillary case (with the 2 opposing DNA software programs). If a person cannot be convicted based solely on a one-witness treason or perjury accusation, or solely based on a confession or accomplice accusation, perhaps people should not be convicted based solely on the output of one algorithm. The problems with this sort of requirement might include the arbitrary nature of how many algorithms exist to analyze a particular problem (e.g. if a Google Earth driving estimate is available, do we really need a Mapquest or Apple Maps estimate? Perhaps that is exactly what we need!). Perhaps such a rule would have a salutary effect on the market for algorithms, to ensure that alternatives exist, incentives for open source options are greater, and research is ongoing and robust.

More broadly, these rules at most govern individual cases with individual lawyers, judges, and litigants. A more ambitious project would be to ensure that algorithms are created and designed and implemented in ways that are responsible. As I mentioned earlier, breath-alcohol machines are subject to a number of front-end safeguards, and other algorithms (such as DNA software, but also LLMs, off-the-shelf machine learning classifiers trained on privileged data sets, and other categories of algorithms as they arise) might be subject to similar requirements, including IEEE-approved stress testing, NIST access, research licenses offered to universities as a condition of getting a particular favorable governing rating, and the like.