December 2021

AI for Judges

A Framework

CSET Policy Brief

AUTHORS
James E. Baker
Laurie N. Hobart
Matthew G. Mittelsteadt

Introduction

Artificial intelligence is a pervasive and persistent part of our lives and will become more so in the future. AI is embedded in shopping algorithms, navigational aids, and search engines. Algorithms drive social media – and, increasingly, vehicles. Diverse fields such as human resources, finance, and medicine all rely on AI. AI detects fraud and trades stock.[1] Studies show that certain AI applications identify tumors with greater accuracy than do medical personnel. Generation Z has come of age in the era of algorithms.

As AI is transforming the economy and American society, it will also transform the practice of law and the role of courts in regulating its use. Law firms use AI applications to conduct discovery. At least 75 countries use facial recognition for domestic security and law enforcement purposes.[2] AI is used to determine travel patterns, to link suspects with crime scenes, and to populate watch lists. Between 2011 and 2019, the FBI used its facial recognition algorithm to search federal and state data bases, such as visa and license data bases, over 390,000 times.[3] AI will also directly and indirectly impact the legal fields of administrative law, contracts, and torts. AI, for example, underpins the advent of smart contracts and new forms of contractual due diligence. Likewise, AI presents new questions about old issues of tort responsibility and liability where AI is used to drive vehicles and make medical diagnoses.

Law rarely, if ever, keeps pace with technology. The legislative and appellate processes simply do not move at the same pace as technological change, and could not do so if they tried. Likewise, scholars and commentators are currently better at asking questions than answering them. As AI applications and cases make their way to court, however, judges do not have the luxury of waiting for answers. As AI applications and cases arise in litigation, judges will confront novel issue after issue. The common law of AI cannot wait. This report is intended to provide a framework for judges to address AI.

In particular, the report considers how AI will impact courts by addressing two question sets.

(1) What role should, will, or might judges play in addressing the use of AI in American society? And, relatedly, how will AI and machine learning impact judicial practice in federal and state courts?

The first section of this paper addresses these questions by considering three purposes of law as well as three judicial roles: judges as evidentiary gatekeepers; judges as constitutional guardians; and judges as AI consumers.

(2) Having identified these roles, what do courts need to know about AI to effectively adjudicate its use by litigants and make informed decisions about whether to use AI as a judicial tool?

The second section addresses these questions by highlighting technical aspects of AI that are likely to play a central role in how AI is adjudicated in courts.

This report is not intended to identify and answer every question that AI might present in a court. There are too many questions to answer. Rather, the goal is to identify *some* of the questions and challenges with the purpose of:

- encouraging judicial inquiry into AI, including areas of likely litigation focus;
- identifying aspects of AI that should inform how judges shape their decisions and avoid unintended case law effects; and
- suggesting a framework for addressing AI in court.

It is for judges to develop a common law of AI. This report is intended as a place to start the intellectual journey ahead.

The Role of Judges and Courts – Framing the Challenge

AI in Less Than a Nutshell

Artificial intelligence has been described as ungovernable. It might be if one were to try and regulate the field in a singular manner, that is, with a single law or case. (Imagine a scenario, for example,

where *Carpenter v. United States*,[4] the Supreme Court's 2018 case applying the Fourth Amendment to cell site location information (CSLI), applied in all circumstances involving data aggregation, link analysis, and AI, and not just CSLI.) However, "AI is not a single piece of hardware or software, but rather, a constellation of technologies that gives a computer system the ability to solve problems and to perform tasks that would otherwise require human intelligence."[5] That means law and regulation will need to address multiple scenarios, applications, and technologies.

Specialists refer to three types of AI. Narrow AI, the AI of today, is AI that can perform singular tasks in an optimal manner or near optimal manner. Strong AI, or Artificial General Intelligence (AGI), is AI that can perform multiple tasks at least at human capacity and move seamlessly from task to task. Super Intelligence (SI) is a notional period when AI is generally smarter than humans. SI raises at least the theoretical possibility that, as suggested in science fiction and Ray Kurzweil's concept of singularity, humans and machines merge into one; or, less optimistically, that machines control humans or perhaps even turn to humans as a source of carbon energy, as Nick Bostrom's thought experiment "the paperclip maximizer" suggests. Experts debate whether and when AI will move from its current narrow iteration to AGI. AI philosophers worry about whether we will get to SI. For sure, SI is the stuff of science fiction, but it is of judicial importance because it is this category of AI that tends to dominate the public impressions of AI, and thus jury pool impressions about AI.

AI has been around in concept at least since 1950, when the English computer scientist Alan Turing published his article, "Computing Machinery and Intelligence." However, six related technological developments have propelled its exponential growth in the past two decades: burgeoning computational capacity, cloud computing, sensors, big data, algorithms, and machine learning. Algorithms are mathematical formulas that guide software. "Machine learning" describes the different ways that software driven machines can be trained to "learn," to perform tasks and improve function. Deep learning is one method for teaching machines to learn. It relies on neural networks, which some

commentators equate to the human brain's neural networks, although the two are quite different. An input is broken down in numerous internal and hidden layers within the network. The problem for specialists is that while a human-designed algorithm drives the neural network, the parameters, weights, and calculations conducted within the internal neural network are not always transparent or understandable to humans. The hidden processes within networks are sometimes referred to as the "Black Box" of machine learning.

Judges need to understand machine learning and deep learning for three reasons:

- The adjudication of AI will necessarily also entail inquiry into, and adjudication of, AI's related technologies.
- Most machine learning is iterative and ongoing with algorithms adjusting formulas and accuracy as they encounter new data. Thus, some AI applications may need to be re-litigated on an ongoing basis even where there is threshold precedent addressing whether the same AI application may be admitted into evidence.
- Many of the pending due process issues associated with AI derive from the nature of neural networks and "black box" aspects of deep learning and machine learning generally.

Fortunately, the role of judges is a narrow one, and there are several straightforward ways to structure the analysis. One way to do so, for example, is to consider the purposes of law and to ask with each AI application:

(1) Does the public or private entity in question have the authority to act as it has, and has it observed any applicable limiting boundaries;

(2) What process is required or due and has it been followed; and

(3) Is the AI's use consistent with our constitutional values?

Another way to frame judicial inquiry into AI is to consider the different roles judges play, or will play, with respect to AI as evidentiary gatekeepers, constitutional guardians, and potential consumers of AI themselves.

Three Purposes of Law

Law serves three purposes. First, it provides authority for public and private entities to act while placing boundaries around those actions. With AI, for example, statutes like the Stored Communications Act and the Privacy Act place requirements on when the government can access data and for what purposes. Section 230 of the Communications Decency Act delimits responsibility for what is posted on internet service providers and social media, a process made more complex by algorithms automatically driving content to consumers at machine speed. However, these laws were passed before the age of AI. For courts, this means interpreting and applying old law in new contexts not contemplated at the time the legislation was passed or case law developed.

Second, law provides essential process. Indeed, with AI, process is essential. The development of AI is exponential where the development of law is linear. Case law never keeps up with "Moore's law" (a prediction by Intel executive Gordon Moore that the processing capacity of transistors on a microchip would double approximately every two years). That means legislation or case law rarely anticipates or addresses every issue that will arise in a substantive manner. But law, and case law, can always define the process by which substantive issues are addressed and with what measure of accountability.

Most AI applications present some form of what we refer to as the "centaur's dilemma." The centaur imagery is drawn from Defense Department vernacular for human-machine teaming, which finds its roots in the mythical creature that is part human and part horse. With many AI applications, a central legal, ethical, and policy question is to what extent will the AI act at human direction versus in an autonomous manner. Is AI augmenting human decision, informing it, or supplanting it? The centaur's dilemma presents a

process question essentially asking where in the AI loop human control will be, can be, should be, or must be asserted. It is a "dilemma" because the more human control that is asserted at the time of use, the less likely the operator will reap the full benefits of the AI's capacity to find meaning in data and do so at machine speed—or in the case of cyber conflict, respond or defend against an adversary's machine speed attacks. Law and courts provide one mechanism to regulate AI so that it enhances rather than undermines the quality of human decision-making or human-machine decision-making.

For courts, process equates to judges serving as gatekeepers. For example, courts might determine prospectively, through the issuance of warrants and orders, or retrospectively, on motions to suppress, when and how AI is used as an investigative tool when AI outputs might serve as probable cause predicate for investigation. Or courts might determine, through application of the Federal Rules of Evidence and their state equivalents, when AI-generated outputs can serve as evidence. Courts must determine whether the human and the machine acted as intended or required by law and did so in an accurate manner.

Third, the law expresses legal and societal values, such as those values found in the First, Fourth, Fifth, Sixth, and Fourteenth Amendments to the Constitution. AI presents myriad new contexts to test these values. What process is due, for example, when an algorithm informs or makes a decision regarding access to government benefits? May the government search open-source data with AI algorithms without probable cause if doing so might chill speech or associational activities? Courts may choose to treat "AI" collectively as a single new technology, as the Supreme Court did in 2001 in *Kyllo v. United States* with respect to "thermal-imaging devices," concluding that the use of a thermal imaging device to detect an external heat signature from a home constituted a Fourth Amendment search.[6] Or courts may address AI as a constellation of technologies, reviewing each application or component on a case-by-case basis. Cases like *Carpenter* suggest that Courts understand that there is something different about the capacity to aggregate data, find meaning in data that humans

cannot see, and do so retroactively and at machine speed that will test and evade contemporary constitutional doctrines and values in new ways. And *Carpenter* was a case about cell phone towers, not AI!

Conscious of these three purposes of law, judges describe their roles in different ways. Often judges refer to themselves as evidentiary gatekeepers as well as constitutional guardians. With AI, however, judges may serve not only as gatekeepers and guardians but also as AI consumers. Let's consider each of these roles in AI context creating a framework for applying law to AI in court.

Judges as Gatekeepers

Judges serve as gatekeepers in multiple ways. This starts with the role of judges in approving search and arrest warrants that may rely on AI-generated outputs and insights as probable cause predicates. Where the government has not relied on a warrant, judges perform this role retroactively by ruling on motions to suppress evidence. Judges as evidentiary gatekeepers will also need to determine whether and when AI evidence will assist the fact finder and is admissible in court. The Federal Rules of Evidence, and their state equivalents, will help guide this determination. The Supreme Court's *Daubert*,[7] *Crawford*,[8] and *Carpenter* cases may also inform the evidentiary questions presented. However, these cases and rules were not written with AI in mind. And, currently, there are few federal or state cases or jury instructions that address AI. Judges will, of course, interpret and apply these cases and rules to AI in the specific contexts presented and do so consistent with the law of the jurisdiction in which they practice.

Federal Rules of Evidence 401-403, 702, 902(13) & (14)

Under Federal Rule of Evidence 401, evidence is relevant if "(a) it has any tendency to make a fact more or less probable than it would be without the evidence; and (b) the fact is of consequence in determining the action."[9] Rule 402 states that relevant evidence is admissible unless the Constitution, a federal statute, the other

Federal Rules of Evidence, or other rules prescribed by the Supreme Court apply and would exclude the evidence.[10] Due process or confrontation clause concerns, for example, might bar certain AI evidence from admission. Rule 403 allows a court to exclude relevant evidence if its probative value is substantially outweighed by a danger of unfair prejudice, confusing the issues, misleading the jury, undue delay, wasting time, or needlessly presenting evidence.[11]

Many of the threshold evidentiary issues associated with AI will likely be litigated under the rubric of Rules 402 and 403 or their state equivalents. That is because relevancy in most or all jurisdictions is broadly defined, and most AI applications are essentially tools for assessing probability—in theory, making them inherently relevant in assessing whether something is "more or less probable." The issue is one of reliability generally, and of appropriate use in the context presented, specifically.

Rules 402 and 403 are essential because the evidentiary use of AI will invariably present questions about discovery and due process, such as whether there is a right to access an underlying algorithm or data when it is used to generate evidence or inform judicial decisions. Further, there is a risk that litigation over AI will present the figurative trial within a trial and risk confusing the jury under Rule 403. In addition, courts might apply Rule 403 to exclude AI evidence that is biased or otherwise unreliable. Inquiry is prudent; otherwise, juries may assume AI evidence has the imprimatur of "science" or "technology" in the context presented, potentially lending it false authority or undue weight, or permitting its use in a manner for which it was not intended.[12]

If relevant and material, judges will also need to decide in what manner and to what extent to require authentication of the AI evidence offered and how, if at all, to validate its reliability. This will bring Federal Rules of Evidence 702 and 902 into play, as well as *Daubert* and *Crawford*. Rule 702 governs the admissibility of expert witness testimony. It provides:

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

Rule 902 covers self-authenticating evidence, such as official records, and newspapers. In 2017, subparagraphs (13) & (14) were added to Rule 902 to address, among other things, the admission of digital evidence and machine-generated records, which in theory now are self-authenticating.

> *(13) Certified Records Generated by an Electronic Process or System.* 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). The proponent must also meet the notice requirements of Rule 902(11).
>
> *(14) Certified Data Copied from an Electronic Device, Storage Medium, or File.* Data copied from an electronic device, storage medium, or file, if authenticated by a process of digital identification, as shown by a certification of a qualified person that complies with the certification requirements of Rule (902(11) or (12). The proponent also must meet the notice requirements of Rule 902 (11).

These rules cover digital photographs and other digital documents as well as data "generated by an electronic process or system," in

other words, material like AI-generated outputs and data. This might, for example, include the output from an AI-driven radiological machine, output from a hiring algorithm used to sort through job applicants, or the program history of a driverless vehicle. Judges will need to decide whether AI-generated outputs qualify for admission under FRE 902(13), and if so, whether particular AI applications produce "accurate results."

Both artificial intelligence and the interpretation of AI outputs are complex. Courts will have to determine the appropriate means to verify AI outputs. This might involve expert testimony, or it might be done through technical means, such as watermarks embedded in an image at the time it is created. Courts will need to determine who is qualified to testify on the accuracy of an AI application. On this question alone, there are many options including: the software engineer, the design engineer, the data engineer, and the company CEO. Courts will need to determine whether the "custodian of records," without more background, is in fact the competent individual to authenticate evidence derived from AI.

Crawford may also come into play. This is the 2004 Supreme Court case holding that in certain contexts documentary evidence should no longer be considered a business record when used as criminal evidence at trial, but rather as testimony for the purpose of triggering the Sixth Amendment right of cross-examination. The Sixth Amendment provides that "[i]n all criminal prosecutions, the accused shall enjoy the right … to be confronted with the witnesses against him." This right is understood to encompass the right to cross-examine witnesses at trial. An algorithm is not a witness. But in *Crawford*, the Supreme Court held that the right to cross-examine witnesses extends, in some cases, to certain out of court "statements" introduced at trial, including statements to the police (as in *Crawford*), as well as "statements that were made under circumstances which would lead an objective witness reasonably to believe that the statement would be available for use at a later trial."[13] Significantly, the Court subsequently held that certain lab reports were testimonial and thus the technician or scientist who compiled the report was subject to examination. Before *Crawford*, many of these statements were admitted into evidence as business

records or under generally recognized exceptions to the hearsay rules. In the absence of clarifying guidance from the Supreme Court, lower courts have struggled to apply *Crawford* to documentary data and other information later introduced as criminal evidence, like lab reports and photographs. That is to say, *Crawford* is applied in an inconsistent, case-by-case manner.

Where AI data is used as evidence in a criminal trial against an accused, the defendant may seek to assert a Sixth Amendment right to question the author of the algorithm. AI-generated information later used as evidence is fertile ground for *Crawford* challenge, including litigation over just who or what is "bearing witness." The software, the learning algorithm, and the computer scientist are all candidates.

Whether *Crawford* is applicable or not, some scholars and practitioners argue that litigants should be able to impeach machines at trial, just as they would human witnesses.[14] The argument is rooted in the Sixth Amendment for sure, but more generally it is rooted in uncertainty about the accuracy of AI-driven machines. Judges, and if not judges then legislators, one scholar argues, "should allow the impeachment of machines by inconsistency and incapacity, and the like, as well as by evidence of bias or bad character in human progenitors."[15] Whether required by *Crawford* or not, legislators and judicial rulemaking bodies might require live testimony "for human designers, inputters, or operators in certain cases where testimony is necessary to scrutinize the accuracy of inputs."[16] Of course, judges might already allow such a process by applying the existing Rules of Evidence, as well as due process. The public policy question here is whether the law or the Rules should require such inquiry, or whether it should be left to the discretion of individual judges to determine.[17]

Daubert, and in certain state systems its predecessor *Frye v. United States*, govern the admission of expert testimony based on scientific methodology, identifying key factors to consider and weigh in the case of *Daubert* and using a "general acceptance" standard in the case of *Frye*. *Frye* is likely more complicated, asking judges to determine when a scientific method "is sufficiently

established to have gained general acceptance in the particular field to which it belongs."[18] In theory, this will entail examination of the specific algorithm and use in question, but also identification of the relevant field of acceptance and what acceptance means for something like "facial recognition," or "predicting behavior."

One way to conceptualize AI evidence is to apply the non-exhaustive list of factors the Supreme Court developed in *Daubert* to determine whether expert testimony based on a specific scientific methodology should be admitted.[19] However, before addressing the factors, judges likely would also need to ask a threshold question: Is the AI application in question "scientific" if it involves no more than coded math equations, or ways to interpret and structure data. Courts might then look to some or all the *Daubert* factors to determine whether the methodology is valid. These include:

(1) Whether the theory or technique in question can be and has been tested;

(2) Whether it has been subjected to peer review and publication;

(3) Its known or potential error rate;

(4) The existence and maintenance of standards controlling its operation;

(5) Whether it has attracted widespread acceptance within a relevant scientific community. [20]

With AI, *Daubert* questions abound. In the case of AI, these factors should be applied to individual algorithms, rather than "AI" as a science or methodology generally. The first question presented by *Daubert* might be directed to identifying the theory or technique or component that is subject to evaluation. Is it:

- the sensors that fed data to the AI system;
- the algorithm;
- the math behind the algorithm;
- the dataset used to train the algorithm;

- the training methodology;
- or, is it the system as an integrated whole that is subject to review?

The second question is: What test is appropriate? And where does one find a baseline against which to establish accuracy? For example, medical diagnostic AI may be compared to physician-diagnosed outcomes. By contrast, it is unclear that an algorithm intended to predict future behavior, such as a criminal assessment tool, can be tested with the same degree of scientific or evidence-based meaning, given the weight such algorithms place on social factors. Human circumstances are endlessly complex, creating multiple influences on behavior; moreover, circumstances do not necessarily determine behavior. In short, predictive algorithms in the criminal context are especially difficult to test, peer review, and assess for accuracy and error rates. For example, there is no way to verify, after an individual has been jailed or sentenced, how that individual's future behavior is affected by the imprisonment. The experience of imprisonment itself, as well as the presence or absence of loved ones outside, might turn a person toward or away from future crime, making it difficult or impossible to verify whether the machine was correct or incorrect in its prediction.

In addition to requiring appropriate peer review, judges will also need to ask the right questions to determine whether "error rates" are accurate and meaningful:

- For example, will, or might the error rates vary depending on whether the AI is tested and reviewed using the relevant local population (database) to which the AI will be applied, as opposed to a national population, or perhaps a more idealized lab database.[21]
- AI imposes maintenance obligations: what dataset is used; is that dataset updated appropriately; is the machine learning monitored by continued testing against known results to ensure the machine is not learning bad habits?
- Courts will also need to determine what peer review means in the context of AI as well as what widespread

> acceptance within the relevant scientific community means.

Each of these questions is compounded by the challenge of peer reviewing and testing the accuracy of algorithms and datasets that comprise the intellectual property and value proposition behind many AI companies. The more challenging question may be: how does one conduct a peer review of a proprietary algorithm or an iterative or evolving machine learning algorithm? Unless courts can demonstratively protect such trade secrets, it may prove hard for courts to apply the *Daubert* factors to many or most AI applications in open court.

One place to start is with the general authority of courts to oversee the admission of evidence, enforce their rulings, and seal records. Another place to start is the Defend Trade Secrets Act, 18 U.S.C. §1835, which in 1996 specifically directed federal courts to protect trade secrets in proceedings arising under Title 18 of the United States Code. Specifically, the section states:

> "In any prosecution or other proceeding under this chapter, the court shall enter such orders and take such other action as may be necessary and appropriate to preserve the confidentiality of trade secrets, consistent with the requirements of the Federal Rules of Criminal and Civil Procedure, the Federal Rules of Evidence, and all other applicable laws."

In context, specific statutes also provide IP protection for AI, such as those found in §705 of the Defense Production Act, which allow the President in the first instance, and courts in the second instance, to use the power of contempt and jurisdiction found in §706 to protect IP relevant to a DPA enforcement or defend against DPA actions.

It is intuitive, but worth remembering that proponents of AI-generated evidence will seek to simplify its admission by limiting or eliminating as many foundational requirements for its admission as possible. Opponents of admission, no doubt, will seek to undermine its relevance and reliability in general, or the purpose for

which it is offered. They will also seek to challenge its relevance and accuracy by seeking access to the underlying algorithm, the data on which it was trained, validated, and tested, as well as what occurs and is weighted inside any machine-learning black box. Thus, courts could face layered adjudicative challenges each time AI generated evidence is offered.

Where AI outputs are admitted, opponents will seek to cross-examine the software engineers responsible for its design. Moreover, because each AI application is different, i.e., it will:

- Have different output purposes;
- Rely on different algorithms;
- Use different machine learning methodologies; and
- Train, test, and validate using different data.

These issues are generally not subject to resolution through the application of case law precedent in the same way, for example, that DNA analysis is now generally accepted in court. One should expect the adjudication of each application and in each context for which the application is offered as evidence.

Judges as Constitutional Guardians

Many judges perceive their roles as one of serving as constitutional guardians. As Justice Jackson said in *Youngstown Sheet & Tube Co. v. Sawyer*, judges should be last and not first to surrender our constitutional institutions and rights. AI will present myriad opportunities to test judicial understanding of constitutional values and their limits. Here are some examples.

First Amendment

AI implicates the five rights protected by the First Amendment: freedom of the press, speech, religion, and assembly, and the right to petition the government. Every time the government, in law or practice, takes an action that can be construed as impeding, restricting, chilling, or favoring one voice or view over another,

there is space for a First Amendment challenge. An inventor seeking a patent, for example, might assert that the government is chilling speech by preventing the inventor from talking about the invention under the Invention Secrecy Act. Consider the myriad issues as well that might arise if the government sought to review and regulate social media postings for foreign interference or undertook to validate the authenticity of information.

Think, too, of the effect of constant or "near perfect surveillance" on First Amendment freedoms, using the Court's phrase from *Carpenter*. Facial recognition, an AI application, is already in use, and many cities use security cameras extensively. Real-time video surveillance machines are able to make predictive identity matches based on photo-memories no human mind could ever catalogue.[22] It is easy to imagine the chilling effect AI surveillance may have on an individual's willingness to speak freely in public, to associate and assemble with politically or religiously motivated groups, or to worship freely.[23] If feeling unimaginative about the potential for AI to enable persistent and pervasive surveillance, one need only Google China's social credit system.

Where AI is concerned, the most contentious First Amendment debates may involve the threshold for initiating investigation of criminal conduct involving domestic terrorism. It is an AI issue because it is likely AI-driven search algorithms and tools that will identify potential threats in the first instance. That means that executive branch lawyers and subsequently courts will need to address the way First Amendment principles are embedded in code. They will need to consider whether First Amendment "constraint" occurs when an algorithm identifies a posting of interest or when a law enforcement officer first looks at the posting and determines whether it meets the threshold for investigation. The FBI *Domestic Investigations and Operations Guide* states, "…investigative activity may not be based solely on the exercise of rights guaranteed by the First Amendment…"[24] One pending question is: what constitutes a sufficient predicate beyond "solely First Amendment activities" to initiate investigation? *Brandenburg v. Ohio* provides a partial answer.[25] First Amendment principles, the Supreme Court concluded, "do not permit a State to forbid or

proscribe advocacy of the use of force or of law violation except where such advocacy is directed to inciting or producing imminent lawless action *and* is likely to incite or produce such action" (emphasis added).[26] But *Brandenburg*, a case involving twelve KKK members standing in a field expressing the possibility of violence, who were filmed and broadcast by invited journalists, was decided in 1969, well before the advent of social media and machine-speed search algorithms. No doubt, courts will be asked to address these predicates in the context of issuing warrants and criminal prosecutions and First and Fourth Amendment motions to suppress evidence.

Deepfakes present another AI area that might implicate First Amendment freedoms. The capacity of AI to convert symbolic language (code) into natural language text along with AI's capacity to discern, recognize, and formulate patterns at the pixel level makes AI a tool of choice not only to identify voices and pictures, but also to mimic voices and alter images. Moreover, this can be done with real-life precision with images or recordings known as deepfakes. Hollywood has, of course, long known about special effects. What makes deepfakes noteworthy for courts is not only the lifelike quality of the technology, but also the accessibility of this capability to the general population. There are tools readily available on the Internet that allow non-specialists to alter photographs and mimic speech with lifelike realism.

This has at least two important manifestations for courts. First, as is often the case, the capacity found its first public manifestation with pornography and pornographic revenge, with digital editors grafting one person's face onto another person's body. In contrast to some areas of AI, some state legislatures have sought to regulate deepfake pornography through criminal sanction.[27] Thus, state courts, but also federal courts in the context of the Child Pornography Prevention Act (CPPA), will likely see an increasing use of AI to generate fantasy porn, revenge porn, and child porn. The questions for courts will include: Is it criminal? And does it fall under some rubric of First Amendment protection? This latter question will bring into play all the ambiguity and vagaries of

Ashcroft v. Free Speech Coalition,[28] the Court's lead case applying the First Amendment to the CPPA.

Second, the same capabilities that allow the lifelike generation of pornography will allow the equally lifelike generation or alteration of evidence. Therefore, judges in their capacity as evidentiary gatekeepers will need to engage in new areas of inquiry and debate involving the authentication of evidence.

Fourth Amendment

The constitutional impact of AI may be most evident and significant for courts in Fourth Amendment context. There is also fairly extensive case law to apply to AI, at least, as a point of departure. The Amendment states:

> The right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated, and no Warrants shall issue, but upon probable cause, supported by Oath or affirmation, and particularly describing the place to be searched, and the persons or things to be seized.

AI-enabled data aggregation and data mining, link analysis, cameras, drones, facial recognition, etc., have the potential to create a system of what Chief Justice Roberts referred to in *Carpenter* as "near perfect surveillance" with respect to CSLI.[29] Such surveillance is possible not only in public spaces but also in our homes and offices,[30] via our phones, computers, and the Internet of Things – connected personal electronic assistants, refrigerators, doorbells, and more.

Of course, AI is not the first technological development to pose Fourth Amendment questions. Courts have long wrestled with how to square the invasive aspects of new technologies with the protection the Fourth Amendment affords against government incursions on privacy. Fourth Amendment analysis about modern technologies has turned on whether the conduct in question constitutes a search, which courts have generally determined by

applying the reasonable expectation of privacy test and its carve-out, the third-party doctrine.

The reasonable expectation of privacy test emerged from the 1967 *Katz v. United States* decision. In determining that police needed a warrant to wiretap a public phone booth, the Supreme Court had to extrapolate beyond the Framers' points of reference of "persons, houses, papers, and effects." The Court held that warrantless wiretaps were unreasonable, reasoning that the Fourth Amendment protects "people, not places."[31] In a concurring opinion, Justice Harlan authored the reasonable expectation of privacy test still in use today.[32] That test considers whether an individual has a subjective expectation of privacy that society also recognizes to be objectively reasonable; if so, that interest is constitutionally protected and any government intrusion on it is presumptively unreasonable in the absence of a warrant.[33]

The reasonable expectation test has the advantage of being capacious and dynamic as technology improves.[34] Arguably, however, it is not very protective as society's expectations of privacy dwindle in the age of social media and the Internet of Things. That is in part because, in two 1970s cases, *Smith v. Maryland* and *United States v. Miller*, the Court created the third-party doctrine. The doctrine posits that if someone voluntarily shares information with a third party, he loses any objectively reasonable expectation of its privacy, and "assumes the risk" the third party may share that information with the government.[35] *Miller* held that law enforcement's acquisition of financial information conveyed by a bank depositor to his bank was not a search within the meaning of the Fourth Amendment.[36] *Smith* held that the police's request that a phone company install a pen register at its central office to record the numbers a suspect dialed was likewise not a search for Fourth Amendment purposes.[37] The third-party doctrine draws a distinction between content information, in which one has a reasonable expectation of privacy, and business records. The *Smith* Court argued that "a pen register differs significantly from the listening device employed in *Katz*, for pen registers do not acquire the *contents* of communications."

With AI, judges will have to consider:

(1) What information do we "voluntarily" convey to various service providers (internet providers, wireless providers, home security and assistance device providers, etc.);

(2) Whether courts should treat that information as content requiring a warrant or business records exempt under the third-party doctrine. AI raises the stakes by allowing private actors or the government potentially to compile and analyze that data at tremendous speed and scale, turning raw data into "content;"

(3) And the extent to which courts will allow "retroactive warrants," i.e., the aggregation, collection, and search of stored data potentially going back years, if not decades.

The Supreme Court has considered the Fourth Amendment implications of modern technologies in two broad categories: (1) where the government uses technology to surveil people directly, and (2) where the government obtains data via the third-party doctrine from private actors who have collected it. In the first category, a series of 1980s aerial surveillance cases may be of interest to courts facing questions about AI-enabled drones. Three cases held that certain aerial surveillance by law enforcement from the publicly navigable airspace did not constitute a search within the meaning of the Fourth Amendment. In 1986, the Court decided that defendants did not have a reasonable expectation of privacy that would preclude surveillance of the curtilage of their home by a plane at 1,000 feet altitude (*Ciraolo*)[38] or the open areas of their industrial complex at 1,200 feet (*Dow Chemical*).[39] In 1989, the Court concluded that surveillance of a backyard by helicopter at 400 feet was not a search within the meaning of the Fourth Amendment (*Florida v. Riley*).[40]

These cases may become specifically relevant in the context of domestic drones used by law enforcement (or by private actors whose records law enforcement subpoenas) or generally relevant as courts consider evolving concepts of privacy.[41] Drones may be equipped with AI-enabled operating systems, allowing them to fly

autonomously or semi-autonomously to gather evidence, or AI-enabled sensors, such as facial recognition. Will police need a warrant to use those drones in the publicly navigable airspace above or near a home or business?[42] Concurring in the judgment in *Florida v. Riley*, Justice O'Connor observed that "public use of altitudes lower than 400 feet – particularly public observations from helicopters circling over the curtilage of a home – may be sufficiently rare that police surveillance from such altitudes would violate reasonable expectations of privacy."[43] In a dissenting opinion, Justice Brennan wrote:

> Imagine a helicopter capable of hovering just above an enclosed courtyard or patio without generating any noise, wind, or dust at all – and, for good measure, without posing any threat of injury. Suppose the police employed this miraculous tool to discover not only what crops people were growing in their greenhouses, but also what books they were reading and who their dinner guests were. Suppose, finally, that the FAA regulations remained unchanged, so that the police were undeniably "where they had a right to be."[44]

We need no longer imagine such "miraculous tools."[45] They are here and they are called drones, enabled by AI and new advanced sensors. FAA regulations currently allow for commercial small drone flight below 400 feet, under certain conditions such as the operator keeping the drone in line of sight.[46] (Operators must apply for a waiver for flights over 400 feet.) Law enforcement may fly drones under those same conditions or apply for a waiver for public drone use.[47] Drones are potentially more discreet than manned airplanes and helicopters, able to approach a residence more closely and quietly. And unlike street cameras, they are mobile.

Drones may become even more invasive if using AI facial recognition or gait analysis, or autonomously tailing a suspect.[48] Some states are moving toward warrant requirements for drones while others are not. This is an example of how AI magnifies the privacy implications of technology, complicating law and policy areas we have not yet sorted and resolved.

In a more recent line of cases, the Supreme Court has tended toward requiring a warrant to use modern technology in criminal searches or to search the technology itself. In 2001, in *Kyllo*, the Court held that law enforcement needed a warrant before using a thermal-imaging device to detect heat-prints emanating from a private home, where, the Court noted, the technology the police used was not yet in general public use.[49] One can imagine this holding might be used to argue against police use of AI-enabled technology, at least so long as the relevant AI-application is not in general use.

In 2012, in *United States v. Jones*, the Court applied a trespass theory of the Fourth Amendment (concurring opinions applied a reasonable expectation of privacy theory), in deciding that law enforcement could not attach a GPS tracker to a suspect's vehicle without a warrant.[50] In *Riley v. California*, in 2014, the Court held police could not search a person's cellphone pursuant to the "search incident to arrest" exception to the warrant clause, concluding that a digital search of a cell phone was much more invasive than a physical search of the materials on a person's body.[51]

Presumably, enhancing technologies and searches with AI will only increase the individual privacy interests at stake. But courts will still need to address competing governmental interests potentially achieved by AI on a case-by-case, or AI application-by-application, basis. As the Court caveated in the 2018 *Carpenter*[52] decision, context matters. That context might be the type of information searched, or the government's purpose in searching, such as for criminal law enforcement or national security ends.

The *Carpenter* decision did not involve AI but appears most apt for AI. With that decision, the Supreme Court continued its trend of requiring a warrant to use or search a modern technology. *Carpenter* "declin[ed] to extend" the third-party doctrine to "a new phenomenon: the ability to chronicle a person's past movements through the record of his cell phone signals," or more specifically, 127 days' worth of cell-site location information (CSLI) that the government had subpoenaed from Carpenter's cell service provider.[53] It was not enough for law enforcement to obtain a

court-ordered subpoena, based on the reasonable suspicion and relevancy standard in the Stored Communications Act;[54] rather, law enforcement use of historical CSLI required a warrant based on probable cause.

The Court described CSLI information as being like the GPS tracking of a vehicle in *Jones*, "detailed, encyclopedic, and effortlessly compiled."[55] Chief Justice Roberts, quoting Justice Sotomayor in *Jones*, wrote that "As with GPS information, the time-stamped data provides an intimate window into a person's life, revealing not only his particular movements, but through them his 'familial, political, professional, religious, and sexual associations.' "[56] The Court distinguished the "exhaustive chronicle" and "revealing nature" of information provided by CSLI records from "the limited types of personal information" collected by pen register and in bank records in *Smith* and *Miller*.

The Court noted, too, that most people carry cell phones everywhere with them, and that CSLI records are typically held by wireless carriers for up to five years, suggesting that law enforcement could look back retrospectively. The Court concluded that "[g]iven the unique nature of cell phone location information, the fact that the Government obtained the information from a third party does not overcome Carpenter's claim to Fourth Amendment protection."[57] The Court, in theory, limited its holding, however, to the facts before it:

> Our decision today is a narrow one. We do not express a view on matters not before us: real-time CLSI or "tower dumps" (a download of information on all the devices that connected to a particular cell cite during a particular interval). We do not disturb the application of *Smith* and *Miller* or call into question conventional surveillance techniques and tools, such as security cameras. Nor do we address other business records that might incidentally reveal location information. Further, our opinion does not consider other collection techniques involving foreign affairs or national security.[58]

If the Court thought CSLI "a new phenomenon," it hasn't seen anything yet when it comes to AI. Data aggregation and connection is not "unique" to CSLI; it is an AI feature. How will *Carpenter* apply to new, AI-enabled technologies or applications? It is an open question, though the trend in the last two decades points to the Court favoring a warrant for invasive emerging technologies or technologies capable of collecting aggregate data over time. The Court's excerpt above suggests security cameras are still covered by the plain sight doctrine, but what of security cameras (or drones) with AI-enabled facial recognition? What if those cameras can instantly search their archives for all pictures of a person, creating a historical record across a web of cameras of comings and goings, perhaps for the past five years? At least with respect to CSLI, the Court required a warrant for a retrospective search. The narrowing language notes that the Court does not opine on real time CSLI. Can law enforcement subpoena security cameras in real time and connect them to other AI-enabled databases that combine facial recognition with instant feedback on a person's criminal and financial records? In either instance, retrospective or real time, the plain view captured on camera is no longer so plain.

Carpenter may well signal the beginning of the end of the third-party doctrine.[59] Even outside the criminal context, it may signal implications for the data used in machine learning.[60] If the Supreme Court was nervous about the aggregation of cell tower data in *Carpenter*, in which data was collected pursuant to legislative authorization, imagine the Court's concern if, and when, it looks at data collection and use from machine learning.

Fifth and Fourteenth Amendments: Due Process and Equal Protection

Government use of AI is susceptible to due process and equal protection concerns where, for example, the government uses AI for DNA testing, criminal justice risk assessments, or watch list selection. The black box aspect of machine learning may compound the challenge. The Fifth Amendment's takings clause may also come into play should the federal government seek to regulate

private AI research or invoke the Invention Secrecy Act to prevent the disclosure of private-sector AI inventions in the interest of national security.

Many police departments and courts across the country use algorithmic risk assessments.[6162] Some police departments use risk assessment tools to predict where crime might occur and by whom.[63] Some courts use risk assessments in pre-trial release, probation, and sentencing decisions.[64] Parole boards also use them.[65] Some of these algorithmic risk assessments rely on machine learning,[66] a capacity that will increase with time.

Using these algorithms to make or inform liberty decisions creates potential Fifth or Fourteenth Amendment due process issues. (It may also create Sixth Amendment confrontation clause issues, as discussed.) To name a few:

- May a defendant meaningfully challenge the logic of an algorithm if the source code is kept from him? Is it enough for him to have access only to the inputs and outputs the algorithm processes and generates?
- If the algorithm uses machine learning, and no one, not even the developer, understands its "analysis," can courts ensure due process of law?
- How may courts test algorithms for accuracy, especially when they predict future (i.e., unrealized) human behavior?

Likewise, racial and other biases in or produced by the algorithms may create equal protection issues. The adoption of risk assessment tools has caused controversy in this context.[67] State legislatures or police departments using these tools might seek to replace, improve, or inform judicial decisions with "evidence-based" algorithmic recommendations,[68] and, in some cases, to decrease the incarceration rate by releasing more people pre-trial and on probation.[69] However, critics argue that risk assessment tools not only have, or could have, racially biased results, but also, through the process of machine learning, exacerbate racial inequalities in the criminal justice system.[70] Among other things, risk assessment

tools may rely on historical data that reflects racially biased policing and legal practices. The ideal of "evidence-based" practice may be appealing, but judges will want to determine whether an assessment tool risks or causes disparate treatment under a mantle of "data-driven" objectivity.

AI's capacity to aggregate, sort, search, and analyze large quantities of data make it a useful intelligence tool. The government might seek to use it in generating and timely maintaining watchlists. However, the application of AI to watchlisting raises procedural due process issues under the Fifth Amendment. Depending on the inputs the government might use, it could also raise equal protection and First Amendment issues.

Even without AI in the mix, some courts have found due process violations in the nomination process for various government watchlists and in the government's redress process for individuals denied flight boarding.[71] Courts addressing watchlisting have applied the *Mathews v. Eldridge* three-factor test to decide what process is constitutionally due, balancing:

(1) the private interests that will be affected by the official action;
(2) the risk of an erroneous deprivation of such interest through the procedures used, and the probable value, if any, of additional or substitute procedural safeguards; and
(3) the Government's interests, including the function involved and the fiscal and administrative burdens that the additional or substitute procedural requirement would entail.[72]

Courts have considered with nuance the individual's right to travel, and to be free from incarceration and from the stigma of being denied boarding or watchlisted.[73] Courts have also considered the government's strong national security interests in watchlisting. Where courts have determined that an individual's liberty interest has been infringed, the cases have turned on the second factor, the risk of erroneous error and the probable value of additional or substitute procedural safeguards.[74] No doubt adding AI to the

equation will increase emphasis on that factor and the relative adjudicative transparency, if any, of applicable algorithms.

In *Latif v. Holder*, for example, the District of Oregon held due process required the government to provide the plaintiffs, who had been denied flight boarding, notice whether they were on the No-Fly list and the reasons for their placement on that list.[75] The notice had to be reasonably calculated to permit plaintiffs to submit evidence rebutting the government's reasons for their inclusion on the list.[76] Where AI algorithms are used, it may be difficult or impossible to determine what factors the AI application considered, erroneously or not, when operating within its black box. The executive, or a court reviewing the executive's actions, might consider whether an AI algorithm could document exactly what factors it considered in nominating a person to a watchlist; and, of course, the executive might impose human review of the AI outputs considering the different forms of human, design, and data bias that can undermine the accuracy of AI outputs.

Judges as Consumers

Judges and courts are also, could be, or will be, consumers of AI. This is already happening with the different search engines used for legal research and for recording and searching documents. Some courts have also turned to predictive algorithms to inform judicial decisions about bail, parole, and sentencing. It is this latter category of AI that has generated the most legal concern.

Most algorithms are based on statistical prediction. In this sense, all algorithms are predictive; however, a class of algorithms also seeks to make predictions about future behavior based on past data. This happens all the time. Shopping algorithms seek to use data about prior purchases (past behavior) to predict the predisposition of individuals to make additional purchases (future behavior). Video platforms like YouTube, which states that "more than 500 hours of content are uploaded every minute,"[77] use algorithms that seek to predict additional videos a viewer might like and watch to generate additional and increased views, and thus potentially advertising revenue. It is called a recommendation algorithm, but what it is doing is pushing product to the viewer based on predictions about

future viewer behavior and likes. Some algorithms are understood to be designed to increase viewer attraction and arguably addiction by increasing the depth of what it is the algorithm is predicting the viewer wants to view: e.g., violence, pornography, comedy, etc.

The immediate question for judges is not whether predicting behavior is inherently good or bad, but whether algorithms that seek to do so are accurate and should inform, or even make, judicial decisions. If so, one needs to identify the benefits and risks of such algorithms as well as mechanisms to increase the benefits and mitigate the risks.

Predictive algorithms are used, or might be used, in a variety of judicial and collateral judicial settings. The most commented upon applications are the use of algorithms to predict pre-trial flight risk, and thus help to determine whether and at what amount to set bail, as well as algorithms that calculate the risk of recidivism and therefore make or inform decisions about parole, parole conditions, or sentencing. However, there are many other law enforcement-related ways in which algorithms may impact judicial decisions. These include predictive policing algorithms that look at past data about the time, location, and nature of arrests, to predict when and where future crimes may occur, so as to increase patrol presences in those areas to either deter crime or address it. Such algorithms are not intended to predict individual conduct, but rather to predict an area and a time where crime might occur, including perhaps the characteristics of individual actors within an area, such as registered sex offenders.

The argument in favor of such algorithms is that they can better focus finite police resources on those areas where crime is most likely to occur based on “neutral data” rather than the hunches, perceptions, or potential racial biases of police officers. The rebuttal is at least twofold. First, such algorithms can generate their own reinforcing and circular logic. The algorithm predicts criminal conduct, police patrols are increased, and additional arrests occur, validating the accuracy of the algorithm. Second, the underlying data may not, in fact, be neutral. Such algorithms may have a disproportionate racial and socio-economic impact where they generate increased patrols in poorer neighborhoods with

historically higher recorded crime rates and larger concentrations of minorities. In this way, they may also reflect past police practices and prosecutorial decisions focusing on resource-poor communities and people of color. They may also have intentional racial impact to the extent they use "race" or socio-economic status as predictive factors.

Several generalized arguments for and against the use of predictive algorithms emerge.

On the one hand:

- Predictive algorithms can identify patterns and trends humans cannot see, and do so rapidly, if not instantaneously, thus curtailing additional risk or harm.
- Predictive AI rests on the premise, and some would say reality, that neither judges nor law enforcement personnel can reasonably predict conduct based on judgment and intuition alone. AI is able to process vastly more data and can excel at analyzing it and finding patterns in it.
- In courts, it could add data to judgments about risk assessment that can inform decisions on bail, parole, and length of sentences.[78] Moreover, because the use of AI is data driven, some argue that in theory, a well-designed algorithm will also be neutral or objective in approach. Where a human might be subject to implicit bias or express bias, an algorithm, proponents might argue, just weighs and reports data-driven facts. Of course, each of these presumptions is hotly contested, which is why courts should hear arguments from both sides where such AI tools are concerned.

On the other hand:

- Western law and criminal procedure are premised on individualized suspicion. This means an individual should be investigated or prosecuted based on articulable facts associated with the individual in question, not patterns found in data regarding the past conduct of other

persons who may share one or more characteristics with the individual in question, like geographic location, or past police practices and prosecutorial decisions.

- Invariably such algorithms focus on characteristics that are, at least purportedly, readily discerned and susceptible to data adaptation and recording. This means classifications such as “race,” gender, marital status, family status, address, and education play a disproportionate role in algorithm design and operation. Conversely, in operation or design, algorithms are less likely to include subjective weights like role models and community connections and participation that might also predict behavior, and some would say do so more accurately.
- Where such classifications are used as factors in algorithmic predictions, they are subject to the risk of bias, intended and unintended. And, as discussed in the next section, this bias may not be intentional, but nonetheless infiltrate an application through training data, or the way computer engineers design the weights assigned to factors.
- Some factors do not account for variation or nuance. Many factors that may appear subject to yes and no answers, and thus “neutral” data scoring, may in reality be more complex and fall along a continuum. “Race” is a good example. Even marital status, which appears an objective data point, may present a continuum of contexts ranging from stable to unstable, and happy to sad. Depending on what the algorithm is intended to predict, nuance can make all the difference in outcomes.
- All these factors are compounded if there is an inability to understand or challenge the underlying algorithm. In all such cases, the lack of transparency undermines the ability of judges and litigators to determine: What factors did the algorithm rely on? How were they weighted? And do those factors in fact reflect the case and parties in question? Where they cannot do so, judges will have to determine if the lack of transparency raises due process concerns.

Mitigation. Judges will first need to decide whether to allow or use predictive algorithms. Aware of these general arguments for and against such algorithms, and to maximize the benefit of using predictive algorithms, where and if they are used, and to minimize the risks, judges might consider the following mitigating steps. There is, of course, a big difference between a shopping algorithm and a parole algorithm. Error in one means a purchase is not made; error in the other impacts the liberty interests of the parties involved.

- Judges should clearly state on the record when they have used an algorithm to inform a decision and the way they have done so, to include the extent to which they have relied on the algorithm. Where judges do so on the record, they should receive additional and appropriate appellate deference as is currently done with most evidentiary rulings that are explained on the record as opposed to those decisions made without explication.
- Just as judges require corroborating evidence before a confession is admitted into evidence, they should require corroboration before relying on an algorithm to inform a decision. They should ask whether the algorithm's statistical prediction aligns with the judge's understanding of the facts. If so, how so? And, if not, why not?
- Where judges rely on algorithms to inform decisions, they should give more deference to algorithms that are transparent in their function, where the factors are identified and the methodology of weighting apparent. Where such factors are not discernible or understandable, judges should ask why not. They should further state on the record why they might nonetheless use the algorithm to inform decisions or why they have declined to use the algorithm.
- Judges should consciously and purposefully distinguish between data that is generated based on group characteristics and data that is specific to the individual in question.

- Judges should insist that any AI used by courts include a mechanism to evaluate its accuracy on an ongoing basis, including mechanisms to identify false positive and false negative rates, along with the trends associated with each.
- Judges should know when racial, gender, or other suspect class factors, or inputs that function as proxies for those factors, such as housing and employment status,[79] are incorporated into algorithmic designs and determine on the record why those factors are relevant to the purpose and function of the AI use in question.
- Where AI is used to make judicial decisions, or not used but available, judges should consciously determine whether that choice should be determined by legislative direction or judicial discretion and do so on the record.

AI Takeaways for Judges

Having identified some of the ways in which AI will enter the work of courts, this section considers what it is judges should know and ask about AI as a technology to effectively play their roles as gatekeepers, guardians, and consumers of AI. Given the role bias plays in the accuracy of AI as well as the traditional role of judges in addressing "bias" in court, this section pays particular attention to the different ways in which bias can impact AI as well as mechanisms of judicial mitigation.

There are many different AI methodologies. There are multiple theories and methods for teaching computational machines to learn. These theories are built into the operative algorithms. In addition to deep learning, which uses deep neural networks as described above, other AI methodologies include: (1) evolutionary or genetic algorithms; (2) inductive reasoning; (3) computational game theory; (4) Bayesian statistics; (5) fuzzy logic; (6) hand-coded expert knowledge; and (7) analogical reasoning.[80] Within the category of machine learning (which includes deep learning), there are multiple ways to teach a machine to learn using data. The three most common are supervised learning, unsupervised learning, and reinforcement learning.[81] Computer engineers must also decide

how much depth and breadth (a.k.a. width) to apply to a deep learning neural network. In other words, how many nodes to include with each internal layer (breadth) and how many layers of internal inputs and outputs it will employ before providing an output (depth). With facial recognition, for example, breadth might represent the number of points in an image an algorithm searches. Depth might be illustrated by the number of times the algorithm goes through this process of breaking an image down into discrete nodes before providing an output. However, the greater the depth – the number of layers in the neural network – the harder it will likely become to determine which factors were determinative in the output prediction. This could become important to the extent there is risk or concern that bias, or some other factor might undermine the accuracy of the outcome. This is also why many algorithms are designed to provide outputs (plural), for example, a range of match faces with facial recognition, or a range of products with a shopping recommendation algorithm.

All of this means that with each AI application, as opposed to AI generally, a court will need to satisfy itself that the specific AI application, its design, and its specific use meet the foundational requirements for the purpose for which it is being offered into evidence or used by a court. This point leads to two questions judges should always ask about AI:

> **1)** Does the AI fit the context for which it is offered or used; in legal terms, is it material?

Courts should pay attention to whether a particular AI application is a good "fit"[82] for the purpose for which it is proffered. Some criminal risk assessments, for example, are designed to determine which individuals might benefit from alternatives to incarceration, such as parole, counseling, etc. These algorithms might have less relevance, and reliability, when used to determine sentencing.[83] That will depend on the input factors, the weight assigned to those factors, the data on which the algorithm was trained, and the nature of the confidence thresholds applied to the output. Thus, while intuitive, courts should pause and ask not only whether the AI at issue is relevant and material to the matter before the court,

but for what purpose the AI was specifically designed, and whether the AI output will materially inform the fact finder.

2) Is the use case reliable?

Even when an algorithm is being used for the purpose for which it was designed, there may be data or design reasons why the reliability of the output will decrease in a specific context. An AI algorithm, for example, may have been designed for and tested upon a particular population. However, the population or purpose for which the AI output is being offered in court may be addressed to a different population with substantial differences from the test population, resulting in less accurate outputs than the lab-tested confidence threshold. In the vernacular of the field, this would constitute inappropriate deployment bias. For example, the FBI facial recognition application, which relies on state license data bases among other U.S. data, would not have the same match accuracy if run against a different input demographic, let's say the population of another country. That is not because the algorithm is intentionally biased, but because it has not been trained against a comparative population pool.

With machine learning, there is also risk that a neural network will rely on inapt factors in making its output predictions. Judges will want to know whether this is possible, and regarding which factors, before allowing a jury to assess the weight of AI evidence or before using an algorithm themselves to assess bail or recidivism risk. For example, a judge would want to know if it were possible for an algorithm to weigh an inappropriate factor, such as "race," gender, or religion (or a proxy) as a recidivism factor, either alone as an input, or as is more likely, as one of multiple factors weighed by the neural network.

Humans are always involved. Machines do what they are programmed to do, not because they choose to do so, but because they are programmed to do so. Software drives machines. And, humans, in the first instance, write software and design programs. That means that behind each AI application there are human choices, human values, and human bias that may impact the operation of the algorithm and the accuracy of its results.

Under current AI vernacular, humans are said to be "in-the-loop," "on-the-loop," or "out-of-the-loop," when it comes to AI. As implied, in-the-loop describes humans in functional control of an application, deciding when and how it is specifically used. On-the-loop describes humans observing AI, but not controlling the AI, but with the option to do so. Out-of-the-loop describes an autonomous system operating automatically. However, the terms are imprecise in at least two regards. First, they describe a wide variance of conduct within each category and thus may convey a sense of control and oversight that is, in operation, absent. More to the point, they are insufficiently descriptive to apportion accountability and responsibility for the purpose of legal judgments.

Take the example of a "driverless car." Some "driverless cars" are configured to employ a safety driver as observer or in the case of a semi-driverless car a driver with shared responsibility for the vehicle's operation. Other driverless cars, without a human in the car, may operate under remote human control. In each of these scenarios, at any moment the vehicle may be driving autonomously without human control; following the explicit direction of the remote or present driver; or the human driver may be keenly observing the operation of the vehicle without overriding the car's computers. In thirty seconds on a road, humans could be said to be out of, in, and on the loop.

Second, however described, a human is always involved with an AI application. For courts, the questions will be: Who designed the algorithm? Who trained the algorithm? Using what data? Who collected the data? Validated the data? Who used the algorithm or monitored its use? In turn, this means that where, for example, *Crawford* applies, there are multiple persons who might be called as witnesses regarding the design and operation of an AI algorithm.

It also means that there will be persons who can, if relevant and material, provide answers to the sorts of questions essential to authenticating and validating the use of AI, like:

(1) What is the AI trained to identify and how has it been weighted and how is it currently weighted?

(2) Does the system have a method to transparently identify these answers and if not, why not?
(3) Are the false positive and false negative rates known? How do those rates relate to the case at hand?
(4) How has AI accuracy been validated and is the accuracy of the AI updated on a constant basis?
(5) Is authenticity an issue?
(6) And how do each of these questions and answers align with how the AI application is being used by the court or proffered as evidence?

Judges might also consider that a qualified AI expert ought to be able to credibly answer these questions, or perhaps they should not be qualified as an expert to address the application at issue.

AI predicts; it does not conclude. That is one reason why engineers use the term "confidence threshold" in describing the accuracy of an application. In the case of the Google search algorithm, for example, the algorithm is predicting that one of the provided links will respond to the query. Accuracy may therefore depend on the design of the algorithm and whether it is intended to provide multiple outputs for human actors to assess or singular responses. The Government Accountability Office (GAO) reported the FBI's conclusion that that the FBI facial recognition application had an 86 percent detection rate when an input image was compared to at least fifty potential output matches drawn from state license data bases; the threshold dropped dramatically as the number of output numbers dropped.[84] Judges, therefore, should ask what factors might impact predictive accuracy. One reason different datasets are used to train, test, and validate AI is accuracy has sometimes been shown to wane when AI is applied in "real world" conditions and outside lab settings.

At present, the accuracy of AI depends on the quality *and* volume of the data on which it is trained. If an algorithm has been trained on only one picture, let's say a picture of an African big cat, then it will be less likely to correctly identify or distinguish between a fur coat, a domestic cat, and a cheetah. This is an important limitation on the capacity and accuracy of current AI. Moreover, volume here is not measured in hundreds, but in hundreds of

thousands of images. A human performing the same task with only one picture will more likely identify the picture using intuition, judgment, and experience, as well as contextual or situational factors the algorithm may not be trained to detect, like terrain and backdrop.

The quality of data is also important. Dated data, known as stale data, is more likely to generate inaccurate results. A facial recognition algorithm trained on license pictures or parole pictures is more likely to identify pictures reflecting the demographics represented in the data bases. This has the potential to increase the false negative rate for underrepresented groups or to increase the false positive rate for over-represented groups.

Likewise, data may possess flaws that impact algorithms, but not humans. Algorithms, for example, may discern links in data or perceive patterns in data creating matches, based on elements or numeric formulas that are unintended or that humans would not discern. For example, the algorithm may match numbers and pixels based on irrelevant factors, such as a common backdrop, labeling, or lighting in photos. If this occurs within the neural network, it may skew a result in a manner unseen and unknown to the user. The output is a face, but the user does not know this face has been passed through to the output stage because of similarities in the picture backdrops, not the face itself.

As will be seen below, this makes certain predictive algorithms particularly susceptible to error. It also makes it essential that judges and fact finders understand the way data can embed witting and unwitting bias into algorithmic design impacting predictive accuracy.

The heart of AI is the algorithm. If the accuracy of an AI application often depends on the amount of data on which it is trained, it depends even more on the algorithm that is applied to that data. An algorithm is a mathematical formula that guides the software, determining which data is selected and how it is weighted. One might liken the algorithm to the recipe a chef uses in a kitchen. However, this chef supplements his recipe every time he cooks in a way that only he knows. What is more, the sous

chefs supplement the recipe when no one is looking. Thus, in some cases no one can be quite sure what gives the recipe its distinctive taste; and, if the chef knew, he would not tell because he wants customers to continue to come to his restaurant.

This means that the heart of many disputes about the use of AI in court will revolve around access to and disputes over the accuracy of algorithms. This is the proprietary secret most AI companies want most to protect, because it is the recipe to their market success and because too much inquiry may undermine confidence in the AI's capacity.

Among the questions for judges to contemplate before an AI application is used or admitted into evidence are:

- To what extent will the court allow parties to discover the content of an algorithm? The data on which the algorithm was trained?
- If discovery is permitted, what safeguards, if any, will the court use to protect the proprietary value of the application?
- When, if ever, does due process require access to an underlying algorithm or its supporting data?
- To what extent is such discovery necessary to apply *Daubert* or *Frye*?

All of which might lead to the additional question:

- Will the court or a jury be able to understand the underlying technology, and is such understanding necessary for a fair adjudication of the facts? If so, what mechanism is appropriate to provide that understanding?

"Bias" is inherent in AI and can impact AI accuracy. For lawyers and judges, bias is often associated with the human application of stereotypes or prejudices to an ethnic, gender, racial, or other identity group. In the AI context, such bias is evident, for example, in the Government of China's use of AI tools, like facial recognition, to track, control and discriminate against its minority Muslim Uyghur population. In U.S. law, such categories are generally recognized as "suspect classes" under the equal protection clause

of the Fifth Amendment, as applied to the Federal Government, and the Fourteenth Amendment, as applied to the individual states. Application of law that treats classes of persons differently from the populace as a whole, if challenged in court, must pass either a strict scrutiny test, intermediate test, or rational basis test depending on the class. Racial classifications receive strict scrutiny requiring the government to show: (1) a compelling government interest for the disparate treatment; and (2) that the means used are narrowly tailored to accomplishing the compelling interest. Gender is subject to intermediate scrutiny, in which case the disparate treatment must further an important government interest and do so by means that are substantially related to the interest.

However, when AI specialists refer to algorithmic bias they are generally referring broadly to the amount of variance between an algorithm's output and the desired outcome, not necessarily legal bias of the sort addressed by the equal protection clause and courts. This sort of variance bias can be caused by witting human prejudice, the sort of bias courts typically address; by cognitive bias of the sort behavioral scientists typically address; and by design and data flaws of the sort computer scientists typically address. Unintentional bias is often difficult to discern because it is embedded in an AI system's design or in the data used to train an algorithm. The problem may be aggravated when the algorithm is both human-generated and machine-generated – a centaur – making it all the harder to transparently see where and how cognitive bias might have entered the system. However, where a judge might put her reasoning for a sentence on the record allowing appellate courts and the parties to understand what occurred and why, the algorithm might act within the black box between input and output. Decision-makers may subsequently place undue reliance on AI outputs that do not warrant such reliance because they are predicated on biased input.

While there is a tendency to believe that "numbers are neutral" and present objective truths, they may still produce erroneous results.[85] However, through careful engineering, thoughtful use of data, and by adjusting algorithmic weights, it is possible to create AI systems with lower margins of error.[86] Judges as evidentiary gatekeepers

can mitigate or block the use of weak or biased AI, by courts and in courts, by asking the right foundational questions. This starts with an understanding of the different and related forms algorithmic bias might take. The United Nations Institute for Disarmament Research suggests several categories and sources of algorithmic bias. We explicate and add to that list here.

Statistical Bias: This might occur when an algorithm's predicted outcomes deviate from a statistical standard, such as the actual frequency of real-world outcomes.[87] This can be caused by bad statistical modeling or incorrect or insufficient data, and thus is not, per se, an AI issue so much as a data issue about which judges should aware. The concern is illustrated with the notional example of calculating the infection or mortality rate of a disease, such as COVID-19. Initial modeling of the COVID rate of infection differed widely, in part, because the models could not account for who had the disease in an asymptomatic manner. Thus, it was only within closed data samples, such as the passengers aboard a cruise ship, that the models could account for an asymptomatic pool where all the passengers were tested before they were allowed to leave the vessels. But then there was risk of having too small a sample pool, nor one that was necessarily a random or representative cross-section of the population.

Moral Bias: This occurs when an algorithm's output deviates from accepted norms (regulatory, legal, ethical, social, etc.).[88] For example, an algorithm may weigh factors that the law or society deem inappropriate or do so with a weight that is inappropriate or inapt in the context presented. Artificial intelligence does not understand the world like humans, and unless instructed otherwise, its results can reflect an ignorance of norms found, for example, in the First Amendment and the equal protection and due process clauses.

Training Data Bias: Like humans, AI learns from experience; however, its experience is based exclusively on the electronic data fed to it, often hand-selected by a human developer. Inaccuracies or misrepresentations in this data can perpetuate biases by embedding them in algorithmic code.[89] For example, an algorithm

intended to identify potentially successful job applicants might rely on past successful job performance as an indicator of future successful job performance and derive from that data certain preferred hiring characteristics like school-attended and experience. However, the underlying data might be dated, for example, from a period when women or minorities were not well represented in numbers in the relevant employment market or school admissions criteria. Thus, the algorithm might exclude candidates who might perform even better than the "successful job performer" dataset from the past.

Inappropriate Focus Bias: This occurs when an algorithm's training data is ill-suited for the algorithm's use.[90] Examples of this form of bias are usually illustrated with reference to algorithms that identify factors within neural networks that as a matter of logic are irrelevant to the desired outcome but nonetheless appear to present an accurate outcome. AI sees patterns humans cannot. Thus, an algorithm may match two images based on colors, or backdrops, or lighting, that are irrelevant to the output purpose.

The risk is not just in false positives, the focus of much bias analysis to date, but in false negatives. For example, disparities in the volume of facial recognition data between males and females may lead to higher inaccuracies in identifying female subjects. This has an equal protection and fairness component if it results in an increase in the number of false positives, for example, the number of innocent female travelers identified for extra screening or questioning at airports. But it has security implications if it results in an inability to track and locate known security subjects or threats, for example, a search for a wanted person on CTV camera feeds or an Amber Alert victim.

Inappropriate Deployment Bias: This might occur when a system is used in a context it was not designed for.[91] For instance, a driverless car trained for driving in the United States might not be able to handle driving on the left side of the road in the United Kingdom. A human will adapt to such a change; a driverless car algorithm would need to be trained to do so.

Overfitting: Overfitting occurs when a machine learning methodology is too tailored to the data it has been trained on and does not account for ambiguities or variations.[92] Generally, this problem is solved by ensuring that the data the machine learning algorithm is trained on is separate from the data it will encounter in functional use. The model is then said to be 'generalized' and should be flexible enough to correctly interpret data it has not encountered.[93] If this testing and training data separation is not made, biased results may occur. For example, if a machine learning sentencing algorithm were built on a training set of past offenders, the AI could design its neural network with results custom fit for those specific offenders. If a person reoffended, perhaps with a lesser crime, and his data was used to train the algorithm, there would be a risk that in calculating a sentence the algorithm might find and match his prior personal data and reproduce the sentence from the last time; that sentence would be, statistically, the best match for his case. In essence, the algorithm might conclude, within its black box, "for someone with this background, we give a sentence of X." The effect: sentencing algorithms could have focused biases that target specific individuals. If the same individual had not been included in the training set, the result could be different. In the second scenario, the algorithm would be forced to seek a more generalized result based on the cases of others and would potentially recommend a different sentence.

What is noteworthy here is not that one sentence or the other is correct, or fair. The point is that they are different and the judge who is relying on the algorithm to inform a decision is unaware that the "advice" the judge is receiving is different not because of the individual characteristics of the case and the offender, but because of the way the algorithm was trained. This problem may be especially likely for reoffenders, or for any pool of persons that might find themselves potentially in training data as well as use data.

To mitigate this risk, judges might ask several questions:

(1) Was the current subject of the output prediction in the training set of the algorithm? And,

(2) If so, what steps were taken to ensure the algorithm is not biased against the subject?

(3) How, if at all, has the algorithm been tuned to avoid overfitting? Were industry best practices used to minimize the risk or impact of overfitting?

Indeed, algorithms have been produced that allow engineers to scrub an individual's data from a machine learning algorithm, essentially making it forget that person.[94] Courts or legislatures might also require that an algorithm used to assess risk not include within its training data any individuals to whom the application might be applied.

Outliers. There is also risk in the other direction, where the application match is an outlier. This might occur where the input is sufficiently distinct from the scenarios built into the algorithm's training sets that the algorithm will not know what to do. The situation is analogous to sentencing for a crime that is not included in sentencing guidelines and is not readily analogous to an existing offense. If the algorithm is designed to produce a result regardless of accuracy, it may attempt to force the case into an incorrect box. This could lead to unpredictable, biased, and incorrect results. This might also occur in the case of new crimes that do not fit within the algorithmic model used to evaluate and assess bail or recidivism risk or for which there is an exceedingly small dataset.

Judges might mitigate this risk by asking two questions:

(1) Is the offense or case in question one for which the algorithm has specifically trained, and if so, with what volume of data?

(2) Is the defendant in question an outlier, and if so in what way? And, has the algorithm been designed to account for those ways?

If the answer to either of these questions is no, then there is heightened risk the algorithm will not predict with the accuracy

intended. For sentencing, the threshold for throwing out algorithmic results could reasonably be low, as the alternative, human decision making, is already the standard. In any event, it would seem incumbent on the proponent of using such an algorithm to demonstrate its validity.

Interpretation Bias: This might occur where the output of the algorithm is either confusing or interpreted incorrectly by those working with the technology.[95] One can imagine with certain facial recognition technology that users might expect singular matches, or perfect matches, in contrast to what most facial recognition algorithms do, which is present an array of potential matches and probability of matches, leaving the ultimate interpretation and conclusion about matches to the human user and not the machine. Interpretation bias may also occur because of ambiguity embedded in the algorithmic design. Software designers unaware of cultural or linguistic cues may or may not use phrases and concepts that can skew results. This might also occur when the reasoning behind a match is necessary to understand the value or import of the match.

Unwitting Human Bias, the unintentional infusion into an application of human preferences, stereotypes, values, or knowledge, may also impact AI accuracy. Use of racial and other social identifying descriptors in algorithms is inherently risky. One can imagine how intentional and unintentional human bias might enter the equation if computer scientists embedded what they believed the characteristics of a “race” or ethnicity to be into facial recognition software. "Race” and ethnicity are inherently ambiguous terms covering a wide continuum of individuals. Bias may also occur unwittingly in machine learning applications that may not be designed to rely on social identity descriptors, but nonetheless rely on such characteristics within the neural network black box. In both cases, such bias can lead to both the under and over inclusion of the targeted group, as concepts like race and ethnicity are malleable.

Intentional Bias. Scientists, operators, and decision-makers may use AI facial recognition tools or predictive algorithms to target disfavored or vulnerable groups. Algorithms can be designed to

identify and select certain real and perceived social identity descriptors associated with "race," gender, sexuality, national origin, religion, disability, etc. With this capacity, for example, facial recognition technology can identify and track certain ethnic groups as is the case in China with "Uyghur characteristics." Although clearly pernicious in the profiling of Uyghurs (or more accurately, a band of physical characteristics Chinese state security services associate with Uyghurs), one legal question for judges is when and under what controls are the purposeful use of social identity descriptors appropriate AI search parameters.

Mitigating Bias: There are different ways to mitigate the risk of AI bias. This starts, of course, with the design phase and continues through the testing and deployment phases of AI development. In addition, to the extent feasible, the system's parameters should be known, or retrievable. AI systems should also be subject to a process of ongoing review and adjustment. The rules, if any, regarding the permissible use of social identifying descriptors should also be enunciated and clear.

In court, mitigating bias is a core judicial function and starts with what judges often do best – ask questions, like:

(1) Who designed the algorithm at play, and subject to what process of review?

(2) Are the algorithm's selection criteria known? Iterative? Retrievable in a transparent form? If not, why not?

(3) Does the application rely on a neural network? If so, is there risk that the system will rely on parameters that are unintended or unknown to the designers or operators? Is it possible to identify those potential parameters? How high is the risk? Is the risk demonstrated? How is the risk mitigated?

(4) Is the input query or prompt asking for a judgment, a fact, or a prediction? Is the judgment, fact, or prediction subject to ambiguity in response?

(5) Do the criteria include real or perceived racial, ethnic, gender or other sensitive categories of social identity descriptors, or proxies for those categories? If so, why, and do the criteria survive ethical and constitutional review? Have engineers and lawyers reviewed the way these criteria were weighted in and by the algorithm, as part of the design function and on an ongoing basis? In accord with what process of validation and review?

(6) Are there situational factors or facts in play that might, could, or should alter the accuracy of the algorithm's predictive accuracy?

(7) Is the application one in which nuance and cultural knowledge is essential to determine the accuracy of the AI application, or to properly query the AI application?

(8) Are the search terms and equations objective or ambiguous in character? Can they be more precise and more objective? If not, why not?

(9) What is the application's false positive rate? What is the false negative rate?

(10) Is there disparity shown in the confidence threshold as between classes of persons based on "race," nationality, religion, gender, sexuality, ability, or some other sensitive category? If so, are there logical and objective reasons for such disparity that survive ethical and constitutional review?

(11) What information corroborates or disputes the determination reached by the AI application? Is the application of the AI designed to allow for such real time assessment? If not, is that based on operational necessity, or simply one of design? If not, is there a process for such assessment that occurs after the fact?

Conclusion

As noted at the outset, the common law of AI cannot wait. Judges should move forward applying the framework suggested in this paper along with the knowledge that by asking the right questions anyone who can understand the hearsay exceptions can also understand AI with sufficient capacity to understand when to use AI and when to admit AI outputs into evidence.

With respect to the judicial use of AI:

- Judges might use or forgo AI algorithms when making bail, sentencing, and parole decisions and do so with or without first validating the underlying AI.
- Judges might distinguish between using an AI application to decide and using it to inform a decision.
- Judges also might decide that where an AI application is used to inform or decide questions of liberty – bail, sentencing, and parole – only publicly provided and disclosed AI systems should be used, or only applications that are also transparent to the defendant.

As the law stands today, these are all choices.

With respect to the introduction of AI-generated evidence, courts have even more choice ahead, at least until the applicable rules of evidence change, binding precedent is set, or legislative bodies define a judicial range of choice.

- Judges will have to decide whether to accept statistical assertions alone in validating the use of an algorithm (such as false positive and false negative rates) or require in-person testimony from experts or software engineers before using an algorithm to inform or decide a bail or parole decision or allowing a jury to rely on an AI output as evidence.
- Most judges, we would surmise, will want to ensure that not only the AI algorithm, but also the data, factors, and weighting, are apt for the purpose for which the AI evidence is introduced.

- Judges may also decide that the moving party behind AI evidence bears the burden for demonstrating not only its admissibility but also its validity.

In the immediate future, the most important thing courts can likely do is ask the right questions and put their analysis and application of the answers on record as to whether, why, how, and subject to what evidentiary determinations. We hope this report helps them do so.

Authors

The Honorable James E. Baker, a CSET distinguished fellow, is a professor at Syracuse University College of Law with a courtesy appointment in the Maxwell School. Judge Baker also serves as director of the Institute for Security Policy and Law. He is the author of *The Centaur's Dilemma: National Security Law for the Coming AI Revolution* (Brookings: 2021).

Laurie Hobart is an associate teaching professor at Syracuse University College of Law, where she teaches national security law.

Matthew Mittelsteadt is an artificial intelligence policy fellow for the Institute for Security Policy and Law (SPL) and a guest lecturer for the Syracuse University College of Law.

Acknowledgments

For excellent feedback and assistance, the authors would like to thank Chuck Babington, Keith Bybee, Tobias Gibson, Danny Hague, Matt Mahoney, the Hon. John Sparks, and Lynne Weil; and our research assistants, Thomas Clifford, Thomas Finnigan III, Hannah Gabbard, Rickson Galvez, Alyssa Kozma, and Michael Stoianoff.

Document Identifier: doi: 10.51593/20190019

Endnotes

[1] Darrell M. West and John R. Allen, How AI is Transforming the World, BROOKINGS (Apr. 2018), https://www.brookings.edu/research/how-artificial-intelligence-is-transforming-the-world/.

[2] Stephen Feldstein, The Global Expansion of AI Surveillance, 1 (Sep. 2019), https://carnegieendowment.org/2019/09/17/global-expansion-of-ai-surveillance-pub-79847; National Security Commission on Artificial Intelligence (NSCAI), INTERIM REPORT 12 (Nov. 2019), https://www.nscai.gov/wp-content/uploads/2021/01/NSCAI-Interim-Report-for-Congress_201911.pdf.

[3] U.S. Gov't Accountability Office, GAO-19-579T, *Face Recognition Technology: DOJ and FBI Have Taken Some Actions in Response to GAO Recommendations to Ensure Privacy and Accuracy, But Additional Work Remains* (June 4, 2019).

[4] *Carpenter v. United States*, 138 S. Ct. 2206 (2018).

[5] NSCAI INTERIM REPORT, *supra* note 2, at 8.

[6] *Kyllo v. United States*, 533 U.S. 27, 36 (2001)("While the technology used in the present case was relatively crude, the rule we adopt must take account of more sophisticated systems that are already in use or in development.").

[7] *Daubert v. Merrell Dow Pharm. Inc.*, 509 U.S. 579 (1993).

[8] *Crawford v. Washington*, 541 U.S. 36 (2004).

[9] FED. R. EVID. 401.

[10] FED. R. EVID. 402.

[11] FED. R. EVID. 403.

[12] Andrea Roth, *Machine Testimony*, 126 YALE L.J 1972 (2017) ("Moreover, just as the Framers were concerned that factfinders would be unduly impressed by affidavits' trappings of formality, 'computer[s] can package data in a very enticing manner.' The socially constructed authority of instruments, bordering on fetishism at various points in history, should raise the same concerns raised about affidavits.")(internal citations omitted).

[13] *Crawford v. Washington*, 541 U.S. 36, 52 (2004).

[14] Roth, *supra* note 12.

[15] Id. at 1982.

[16] Id.

[17] Id., at 2031 (discussing other nations' choices).

[18] *Frye v. United States*, 293 F. 1013, 1014 (D.C. Cir. 1923)

[19] *Daubert v. Merrell Dow Pharm. Inc.*, 509 U.S. 579, 579 (1993).

[20] *Id.* at 593–595.

[21] Barabas, Chelsea, Christopher T. Bavitz, Ryan H. Budish, Karthik Dinakar, Cynthia, Dwork, et al. *An Open Letter to the Members of the Massachusetts Legislature Regarding the Adoption of Actuarial Risk Assessment Tools in the Criminal Justice System*, BERKMAN KLEIN CENTER FOR INTERNET & SOCIETY 3 (Nov. 9, 2017), http://nrs.harvard.edu/urn-3:HUL.InstRepos:34372582; *see also* Christopher Bavitz, Sam Bookman, Jonathan Eubank, Kira Hessekiel, and Vivek Krishnamurthy, *Assessing the Assessments: Lessons From Early State Experiences in the Procurement and Implementation of Risk Assessment Tools.* BERKMAN KLEIN CENTER FOR INTERNET & SOCIETY research publication, 7 (Nov. 2018), http://nrs.harvard.edu/urn-3:HUL.InstRepos:37883502.

[22] Gregory Barber & Tom Simonite, Some US Cities Are Moving Into Real-Time Facial Surveillance, Wired (May 17 2019), https://www.wired.com/story/some-us-cities-moving-real-time-facial-surveillance/.

[23] See Jennifer Lynch, *Face Off: Law Enforcement Use of Face Recognition Technology*, ELECTRONIC FRONTIER FOUNDATION 1, 8–10 (Feb. 12, 2018), https://www.eff.org/wp/law-enforcement-use-face-recognition#_idTextAnchor004.

[24] *Domestic Investigations and Operations Guide*, *Federal Bureau of Investigation*, released March 2016, updated September 2016, at 4-4, https://www.justsecurity.org/wp-content/uploads/2019/03/FBI.DIOG_.pdf.

[25] *Brandenburg v. Ohio*, 395 U.S. 444 (1969).

[26] *Id.* at 447 (emphasis added).

[27] Matthew F. Ferraro, *Deepfake Legislation: A Nationwide Survey*, WILMERHALE (2019), https://www.wilmerhale.com/en/insights/client-alerts/20190925-deepfake-legislation-a-nationwide-survey.

[28] *Ashcroft v. The Free Speech Coalition*, 535 U.S. 234 (2002).

[29] *Carpenter v. United States*, 138 S. Ct. 2206, 2210 (2018). *See also* Ryan Calo, *Artificial Intelligence Policy: A Primer and Roadmap*, 51 U.C. DAVIS L. REV. 399, 423 (2017) ("Even assuming away the likely false positives, a reasonable question for law and policy is whether we want to live in a society with perfect enforcement.").

[30] There is a growing body of literature on AI in the workplace; *see, e.g.*, Karen E. C. Levy (2015), *The Contexts of Control: Information, Power, and Truck-Driving Work, The Information Society*, 31:2, 160-74, https://www.tandfonline.com/doi/full/10.1080/01972243.2015.998105.

[31] *Katz v. United States*, 389 U.S. 346, 351 (1967).

[32] *Id.* at 360 (Harlan, J., concurring).

[33] *Id.* at 360–61.

[34] Stephen Dycus, Arthur L. Berney, William Banks, Peter Raven-Hansen, Stephen I. Vladeck, NATIONAL SECURITY LAW, Sixth Ed., Wolters Kluwer (2016) Teachers' Manual, 24-3.

[35] *Smith v. Maryland*, 442 U.S. 735, 744 (1979) (citing United States v. Miller, 425 U.S. 435, 442–444 (1976)).

[36] *Id.* at 744.

[37] *Id.* at 737.

[38] *California v. Ciraolo*, 476 U.S. 207, 215 (1986).

[39] *Dow Chemical Co. v. United States*, 476 U.S. 227, 239 (1986).

[40] *Florida v. Riley*, 488 U.S. 445, 455 (1989).

[41] *See* Troy A. Rule, *Airspace In An Age Of Drones*, 95 B.U. L. REV. 155, 172-74 (2015), and Gregory S. McNeal, *Drones and the Future of Aerial Surveillance*, 84 GEO. WASH. L. REV. 354, 373-83(2016).

[42] *See Kyllo v. United States*, 533 U.S. 27, 33 (2001) (quoting *Dow Chemical*, 476 U.S. at 237, n. 4) ("We have previously reserved judgment as to how much technological enhancement of ordinary perception from such a vantage point, if any, is too much. While we upheld enhanced aerial photography of an industrial complex in *Dow Chemical*, we noted that we found "it important that this

is *not* an area immediately adjacent to a private home, where privacy expectations are most heightened[...]")(emphasis in original).

[43] *Florida v. Riley*, 488 U.S. at 455 (O'Connor, J., concurring); see McNeal, *supra* note 41, at 377.

[44] *Id.* at 462 (Brennan, J., dissenting).

[45] See McNeal, supra note 41, at 383; Rule; *supra* note 41, at 174.

[46] 14 C.F.R. Part 107 – SMALL UNMANNED AIRCRAFT SYSTEMS (May 2021), https://www.ecfr.gov/cgi-bin/text-idx?node=pt14.2.107&rgn=div5.

[47] *Drones in Public Safety: A Guide to Starting Operations*, FEDERAL AVIATION ADMINISTRATION (February 2019), https://www.faa.gov/uas/public_safety_gov/media/Law_Enforcement_Drone_Programs_Brochure.pdf.

[48] Cade Metz, *Police Drones Are Starting to Think for Themselves*, THE NEW YORK TIMES, Dec. 5, 2020, https://www.nytimes.com/2020/12/05/technology/police-drones.html.

[49] *Kyllo v. United States*, 533 U.S. 27, 31–41 (2001).

[50] *United States v. Jones*, 565 U.S. 400, 404-05 (2012).

[51] *Riley v. California*, 573 U.S. 373, 385–98 (2014).

[52] *Carpenter v. United States*, 138 S.Ct. 2206 (2018).

[53] *Id.*

[54] The Stored Communications Act, as amended in 1994, "permits the Government to compel the disclosure of certain telecommunications records when it 'offers specific and articulable facts showing that there are reasonable grounds to believe' that the records sought 'are relevant to an ongoing criminal investigation.'" *Id.* at 2212 (citing 18 U.S.C. 2703(d)).

[55] *Carpenter*, 138 S.Ct. at 2216.

[56] *Id.* at 2217.

[57] *Id.* at 2220.

[58] *Id.*

[59] *See, generally*, James E. Baker, THE CENTAUR'S DILEMMA: NATIONAL SECURITY LAW FOR THE COMING AI REVOLUTION, Brookings (2020).

[60] *Id; see, Patel v. Facebook, Inc.*, 932 F.3d 1264, 1273 (9th Cir. 2019).

[61] Randy Rieland, *Artificial Intelligence is Now Used to Predict Crime. But Is It Biased?* SMITHSONIAN MAGAZINE (Mar. 5, 2018), https://www.smithsonianmag.com/innovation/artificial-intelligence-is-now-used-predict-crime-is-it-biased-180968337/.

[62] *Algorithms in the Criminal Justice System: Pre-Trial Risk Assessment Tools*, ELECTRONIC PRIVACY INFORMATION CENTER, https://epic.org/algorithmic-transparency/crim-justice/

[63] Rieland, *supra* note 61.

[64] *Algorithms in the Criminal Justice System: Pre-Trial Risk Assessment Tools*, *supra* note 62.

[65] *Id.*

[66] Danielle Kehl, Priscilla Guo, & Samuel Kessler, *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing*, RESPONSIVE CMTYS. INITIATIVE (July 2017), https://dash.harvard.edu/bitstream/handle/1/33746041/201707_responsivecommunities_2.pdf?sequence=1&isAllowed=y.

[67] For example, a 2016 ProPublica Study determined that COMPAS was almost twice as likely to falsely identify a black person as a repeat violent offender as it was to falsely identify a white person as a repeat offender. The company contested this finding. Julia Angwin et al., *Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks*, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing. *See also* Sam Davies-Corbett et al., *A computer program used for bail and sentencing decisions was labeled biased against blacks. It's actually not that clear.*, WASH. POST (Oct. 17, 2016), https://www.washingtonpost.com/news/monkey-cage/wp/2016/10/17/can-an-algorithm-be-racist-our-analysis-is-more-cautious-than-propublicas/.

[68] *Loomis v. Wisconsin*, 881 N.W.2d 749, 759 (Wis. 2016), cert. denied, 137 S.Ct. 2290 (2017).

[69] Derek Thompson, *Should We Be Afraid of AI in the Criminal-Justice System?* THE ATLANTIC (June 20, 2019), https://www.theatlantic.com/ideas/archive/2019/06/should-we-be-afraid-of-ai-in-the-criminal-justice-system/592084/.

[70] See, A Letter to the Members of the Criminal Justice Reform Committee of Conference of the Massachusetts Legislature Regarding the Adoption of Actuarial Risk Assessment Tools in the Criminal Justice System (Feb. 9, 2018), https://medium.com/berkman-klein-center/a-lettnener-to-the-members-of-the-criminal-justice-reform-committee-of-conference-of-the-massachusetts-2911d65969df.

[71] *E.g.*, *Ibrahim v. Dep't of Homeland Sec.*, 62 F. Supp. 3d 909 (N.D. Cal. 2014); *Latif v. Holder*, 28 F. Supp. 3d 1134 (D. Or. 2014); *but see Elhady v. Kable*, 993 F.3d 208 (4th Cir. 2021)(reversing district court finding of due process violation, where plaintiffs' travels were delayed but not precluded); *Abdi v. Wray*, 942 F.3d 1019 (10th Cir. 2019); *Beydoun v. Sessions*, 871 F.3d 459 (6th Cir. 2017). For a discussion of the government database at issue in *Elhady*, see Jeffrey Kahn, *Why a Judge's Terrorism Watchlist Ruling is a Game Changer: What Happens Next*, JUST SECURITY (Sept. 9, 2019), https://www.justsecurity.org/66105/elhady-kable-what-happens-next-why-a-judges-terrorism-watchlist-ruling-is-a-game-changer/.

[72] *Mathews v. Elridge*, 424 U.S. 319, 335 (1976).

[73] *Compare Ibrahim*, 62 F. Supp. 3d at 928 *and Latif*, 28 F. Supp. 3d at 1148–51 *with Elhady*, 993 F.3d at 226-7, *Beydoun*, 871 F.3d at 469, *and* Abdi, 942 F.3d at 1033-34 (all determining plaintiffs could not establish the "plus" parts of their "stigma plus" claims because their placement on watchlists did not result in the denial or alteration of any previously held legal right)

[74] See Dycus, et al, *supra* note 34, at 26-6. *But see Elhady*, 993 F.3d at 228 (finding "the weight of the private interests at stake . . .comparatively weak" where plaintiffs' travels were only delayed).

[75] *Latif*, 28 F. SUPP. 3D AT 1162.

[76] *Id.* The *Latif* court left it to the government to fashion the appropriate procedures, but suggested it might provide unclassified summaries or share the classified reasons with cleared counsel.

[77] YouTube Official Blog, https://blog.youtube/press/.

[78] Jason Tashea, *Courts Are Using AI to Sentence Criminals. That Must Stop Now*. WIRED (Apr. 17, 2017), https://www.wired.com/2017/04/courts-using-ai-sentence-criminals-must-stop-now.

[79] Barabas, Chelsea, Christopher T. Bavitz, Ryan H. Budish, Karthik Dinakar, Cynthia, Dwork, et al. *An Open Letter to the Members of the Massachusetts Legislature Regarding the Adoption of Actuarial Risk Assessment Tools in the Criminal Justice System*, BERKMAN KLEIN CENTER FOR INTERNET & SOCIETY 3 (Nov. 9, 2017), http://nrs.harvard.edu/urn-3:HUL.InstRepos:34372582.

[80] United Nations Institute for Disarmament Research (UNIDIR), THE WEAPONIZATION OF INCREASINGLY AUTONOMOUS TECHNOLOGIES: ARTIFICIAL INTELLIGENCE 5 (2018), http://www.unidir.org/files/publications/pdfs/the-weaponization-of-increasingly-autonomous-technologies-artificial-intelligence-en-700.pdf.

[81] *Id.* at 4.

[82] See *Daubert v. Merrell Dow Pharm. Inc.*, 509 U.S. 579, 591-92 (1993).

[83] See Christopher Bavitz, et al, *supra* note 21, *Assessing the Assessments*, at 6-7 (discussing the Wisconsin Supreme Court's warning in *State v. Loomis*, 881 N.W.2d 729 (Wis. 2016) that the risk assessment tool COMPAS was not developed for use at sentencing).

[84] GAO-19-579T, *supra* note 3, at 14.

[85] Joni R. Jackson, *Algorithmic Bias*. 15 J. OF LEADERSHIP, ACCOUNTABILITY & ETHICS 55-65 (2018).

[86] Jake Silberg & James Manyika, *Notes from the AI frontier: Tackling bias in AI*, MCKINSEY GLOBAL INSTITUTE (June 6, 2019), https://www.mckinsey.com/~/media/mckinsey/featured insights/artificial intelligence/tackling bias in artificial intelligence and in humans/mgi-tackling-bias-in-ai-june-2019.ashx.

[87] United Nations Institute for Disarmament Research (UNIDIR), ALGORITHMIC BIAS AND THE WEAPONIZATION OF INCREASINGLY AUTONOMOUS TECHNOLOGIES: ARTIFICIAL INTELLIGENCE 2 (2018), http://www.unidir.ch/files/publications/pdfs/algorithmic-bias-and-the-weaponization-of-increasingly-autonomous-technologies-en-720.pdf.

[88] *Id.*

[89] *Id.* at 2-3.

[90] *Id.* at 4

[91] *Id.*

[92] IBM Cloud Education, *Overfitting*, IBM (Mar. 3, 2021), https://www.ibm.com/cloud/learn/overfitting.

[93] See *id.*

[94] Mathew Hutson, *Researchers Can Make AI Forget You*, IEEE SPECTRUM (Jan. 15, 2020), https://spectrum.ieee.org/tech-talk/computing/software/researchers-can-make-ai-forget-you.

[95] UNIDIR, ALGORITHMIC BIAS, *supra* note 87, at 4.