AI, CAN YOU HEAR ME? PROMOTING PROCEDURAL DUE PROCESS IN GOVERNMENT USE OF ARTIFICIAL INTELLIGENCE TECHNOLOGIES

Chris Chambers Goodman [*]

Cite as: Chris Chambers Goodman, *Ai, Can You Hear Me? Promoting Procedural Due Process in Government Use of Artificial Intelligence Technologies*, 28 RICH. J.L. & TECH. 700 (2022).

---

[*] Straus Research Professor and Professor of Law, Pepperdine Caruso School of Law; Visiting Professor UCLA School of Law; J.D. Stanford Law School; A.B. cum laude, Harvard College. The author wishes to express her profound thanks to Ani Khachatryan for excellent and thorough work researching and editing this article. In appreciation of all the panelists and participants who discussed an early draft at the (in-person!) 12th Annual Constitutional Law Colloquium, hosted by Barry Sullivan and Alexander Tsesis. I dedicate this article to those of us engaged in the scholarly pursuit of equity.

ABSTRACT

This Article explores the constitutional implications of algorithms, machine learning, and Artificial Intelligence (AI) in legal processes and decision-making, particularly under the Due Process Clause. Regarding Judge Henry J. Friendly's procedural due process principles of the U.S. Constitution,[1] decisions produced using AI appear to violate all but one or two of them. For instance, AI systems may provide the right to present evidence and notice of the proposed action, but do not provide any opportunity for meaningful cross-examination, knowledge of opposing evidence, or the true reasoning behind a decision. Notice can also be inadequate or even incomprehensible. This Article analyzes the challenges of complying with procedural due process when employing AI systems, explains constraints on computer-assisted legal decision-making, and evaluates policies for fair AI processes in other jurisdictions, including the European Union (EU) and the United Kingdom (UK). Building on existing literature, it explores the various stages in the AI development process, noting the different points at which bias may occur, thereby undermining procedural due process principles. Furthermore, it discusses the key variables at the heart of AI machine learning models and proposes a framework for responsible AI designs. Finally, this Article concludes with recommendations to promote the interests of justice in the United States as the technology develops.

[1] *See* Henry J. Friendly, *"Some Kind of Hearing,"* 123 U. PA. L. REV. 1267, 1279–95 (1975).

"By far the greatest danger of Artificial Intelligence is that people conclude too early that they understand it.[2]

"Today's AI/ML is uninterpretable, biased, and fragile. When it works, we don't understand why."[3]

"The status quo is not neutral."[4]

I. INTRODUCTION

[1] Public sector resource allocation increasingly relies upon algorithmic decision-making.[5] Algorithms assist governmental actors in allocating benefits, such as Temporary Aid to Needy Families (TANF), Medicaid, and unemployment insurance payments.[6] Automation has increased efficiencies in these processes, and accuracy is an important reason for the increase in the use of machine learning. However, "algorithms still make mistakes, and it is these mistakes that keep legal

[2] Eliezer Yudkowsky, *Artificial Intelligence as a Positive and Negative Factor in Global Risk*, *in* GLOBAL CATASTROPHIC RISKS 308, 308 (Nick Bostrom & Milan M. Ćirković eds., 2008).

[3] James X. Dempsey, *Artificial Intelligence: An Introduction to the Legal, Policy and Ethical Issues*, BERKELEY CTR. FOR L. & TECH. 1, 27–28 (2020) (quoting Arvind Narayanan); *see also id.* at 27–28 (describing the Department of Housing and Urban Development's complaint against Facebook regarding its use of algorithms to deliver housing advertisements and another example of AI making credit decisions that resulted in a disparate impact—even though race was not explicitly considered in the algorithm).

[4] Sandra Wachter et al., *Bias Preservation in Machine Learning: The Legality of Fairness Metrics Under EU Non-Discrimination Law*, 123 W. VA. L. REV. 735, 768 (2021).

[5] *See* Hannah Bloch-Wehba, *Access to Algorithms*, 88 FORDHAM L. REV. 1265, 1273 (2020).

[6] *See* Sarah Valentine, *Impoverished Algorithms: Misguided Governments, Flawed Technologies, and Social Control*, 46 FORDHAM URB. L.J. 364, 385, 413 (2019).

scholars up at night,"[7] leading to substantial costs, particularly for the individuals who are subjected to these mistakes.[8] For this reason, some argue that the use of algorithms is better suited to denying, rather than providing, benefits to needy people.[9]

[2] These benefits enjoy the status of protected "property interests,"[10] and therefore the Due Process Clause requires that the government explain its decisions regarding reductions or terminations of public benefits.[11] These explanations must rely upon "ascertainable standards" in decision-making.[12] By delegating power to AI designers, agencies dilute the

[7] David Lehr & Paul Ohm, *Playing with the Data: What Legal Scholars Should Learn About Machine Learning*, 51 U.C. DAVIS 653, 710 (2017).

[8] *See* Neil Ballantyne, "Not on My Watch!" – A Case Study in the Datafication of Child Welfare in Aotearoa New Zealand 19 (2021) (M.Phil. thesis, Massey University) (on file with author) (explaining that in New Zealand, predictive algorithms designed to assess the risk of children being subject to abuse considered inputs like unemployment and tardiness or missing doctors' and dental appointments as highly correlated with potentially committing child abuse and neglect).

[9] *See* LYDIA X. Z. BROWN ET. AL., CTR. FOR DEMOCRACY & TECH., CHALLENGING THE USE OF ALGORITHM-DRIVEN DECISION-MAKING IN BENEFITS DETERMINATIONS AFFECTING PEOPLE WITH DISABILITIES 4 (Oct. 2020), https://cdt.org/wp-content/uploads/2020/10/2020-10-21-Challenging-the-Use-of-Algorithm-driven-Decision-making-in-Benefits-Determinations-Affecting-People-with-Disabilities.pdf [https://perma.cc/2LS5-L3UF].

[10] Bloch-Wehba, *supra* note 5, at 1272; *see also* Social Security Amendments of 1965, Pub. L. No. 89-97, 79 Stat. 343 (explaining that because of this protection, Medicaid benefits cannot be decreased or withdrawn without compliance of constitutional due process).

[11] *See* Bloch-Wehba, *supra* note 5, at 1275 (explaining that these explanations, under federal law, must present "a clear statement of the specific reasons for supporting the intended action" (citing 42 C.F.R. § 431.210(b) (2019))).

[12] Holmes v. N.Y.C. Hous. Auth., 398 F.2d 262, 265 (2d Cir. 1968) (citing Hornsby v. Allen, 326 F.2d 605, 612 (5th Cir. 1964)); Lightfoot v. District of Columbia, 448 F.3d 392, 400 (D.C. Cir. 2006).

procedural protections for claimants. Congress delegates authority to the agencies, and then the agencies "in essence re-delegate their Congressional authority to computer programmers."[13]

[3] Moreover, some agencies adopt technology even when they know the technology is flawed.[14] One case brought by the American Civil Liberties Union (ACLU) accused an agency of violating due process by depriving disabled Medicare recipients of necessary care.[15] The court found that the agency knew that the software allocated inadequate funding and failed to recalibrate it to allocate the proper amounts.[16] Another case involved a fraud detection system[17] for those applying for unemployment benefits, whereby up to 50,000 people were accused and penalized when subsequent litigation revealed that *the algorithm was wrong 93% of the time*.[18] The state of Michigan, in its efforts to upgrade and modernize its unemployment insurance agency, designed, created, and implemented the Michigan integrated data automated systems, called MiDAS, to flag and

---

[13] Ryan Calo & Danielle K. Citron, *The Automated Administrative State: A Crisis of Legitimacy*, 70 EMORY L.J. 797, 816 (2021).

[14] *See id.* at 804, 819 ("Mounting evidence suggests that agencies are turning to systems in which they hold no expertise, and that foreclose discretion, individuation, and reason-giving almost entirely.").

[15] *Id.* at 823–24.

[16] *See id.* at 824.

[17] *See generally* Andrew Keshner, *What to do if someone else is fraudulently claiming your unemployment benefits*, MARKETWATCH (June 13, 2020, 12:27 PM), https://www.marketwatch.com/story/what-to-do-if-someone-else-is-claiming-your-unemployment-benefits-2020-06-08 [https://perma.cc/L2QS-NJYX] (explaining the basics of "imposter fraud").

[18] Sarah Cwiek, *State review: 93% of state unemployment fraud findings were wrong*, MICH. RADIO (Dec. 16, 2016, 6:03 PM), https://www.michiganradio.org/politics-government/2016-12-16/state-review-93-of-state-unemployment-fraud-findings-were-wrong [https://perma.cc/C8RC-U6C7]; *see* Calo & Citron, *supra* note 13, at 827.

evaluate unemployment eligibility.[19] Almost 30,000 new fraud cases were open, and even though the state was not able to offer evidence to support the fraud allegation in many cases, people were denied benefits.[20] The Sixth Circuit upheld the district court finding violations of due process.[21] Prior to the coronavirus pandemic, California had a fraud detection program that worked well, but the program was discontinued due to its high cost.[22] Other states still use the same system successfully.[23]

[4] So, the question becomes, can we fix these errors? If so, how? There are numerous opportunities in the AI development and deployment process for miscarriages of justice to occur. Because predictive algorithms rely upon data from past group behaviors to predict the likelihood of future individual behaviors, these correlations "necessarily capture[] whatever biases are incorporated into past social behaviors, reinscribing them into predictions

---

[19] Cahoo v. SAS Analytics Inc., 912 F.3d 887, 892 (6th Cir. 2019).

[20] *Id.* at 893–94.

[21] *Id.* at 899–901.

[22] *See* David Manoucheri, *EDD had fraud detection in 2016, then turned it off*, KCRA (Feb. 13, 2021, 9:58 AM), https://www.kcra.com/article/edd-had-fraud-detection-2016-then-turned-it-off/35491386 [https://perma.cc/EWL4-5HJ4] (discussing Pondera Solutions' fraud detection program that California's Employment Development Department used).

[23] Patrick McGreevy, California dropped its guard before it was hit with $2 billion in unemployment fraud, L.A. TIMES (Dec. 21, 2020, 5:00 AM), https://www.latimes.com/california/story/2020-12-21/california-precautions-ignored-employment-development-department [https://perma.cc/77C3-BFLV] (explaining how the Pondera Solutions program, which uses Google technology and publicly available data to flag potentially fraudulent claims, is being used by dozens of federal and state agencies in states like Nevada and Wisconsin).

about future behaviors."[24] For instance, noisy data sets contain too much extraneous information and limit an algorithm's accuracy.[25] The inability to generalize and perform similarly well on training data is another concern.[26]

[5] One of the additional issues brought on by algorithmic decision-making relates to scale. One algorithmic error could simply be one independent error in one case, but it could also be one error in a line of code creating hundreds of thousands of erroneous decisions.[27] It is impossible to know if the error is systemic or individualized without dissecting the processes.[28] Automated systems "create instability and uncertainty that upends people's lives."[29] While some might refer to these outcomes as

---

[24] Dan L. Burk, *Algorithmic Legal Metrics*, 96 NOTRE DAME L. REV. 1147, 1183 (2021) ("[t]hus, the inescapable bias in predictive algorithms stems not from inaccuracies in the date or in the data processing—although they are surely there—but from the distortions inherent in the *entire project* of predictive social correlations. Certainly, the accuracy of data profiling may be a serious and legitimate concern that must be taken into account in assessing proposals such as reliance on algorithmically generated legal metrics.")

[25] Lehr & Ohm, *supra* note 7, at 711–12.

[26] *See id.* at 713–15.

[27] *See* Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1267 (2008); *see also* Cary Coglianese & David Lehr, *Transparency and Algorithmic Governance*, 71 ADMIN. L. REV. 1, 41 (2019) (explaining how procedural due process standards mandate that the government include information concerning an algorithm's accuracy in fulfilling its objectives to ensure that individuals are treated fairly and that procedures are not susceptible to serious error).

[28] Coglianese & Lehr, *supra* note 27 (noting that the presence of an error in an individual's case does not guarantee that there is a systemic design issue since most algorithms produce some errors—without dissecting the process, it is impossible to know if the error is systemic or individualized).

[29] Calo & Citron, *supra* note 13, at 819; *see also id.* at 799, 821–22 (providing the example of an "algorithmic absurdity," in which the creator of the algorithm admitted as a mistake when the Arkansas Department of Human Services used the algorithm which determined that a person whose foot had been amputated had "no foot problems" and thus needed less care).

"errors," in some cases the so-called mistaken outcomes are exactly what the algorithm was designed to achieve—or trained to achieve—or both.

[6] As a result, in state courts, claims challenging denials of Medicaid benefits under due process have increased notably[30] as the Medicaid Act provides a right to a "fair hearing" if benefits are denied.[31] The meaning of a "fair hearing" in the context of a machine-generated decision denying benefits raises important questions: What "evidence" can be presented at such a hearing? By whom may such evidence be presented? Can the finder of fact "understand" what the evidence means? How does one determine whether the decision was "accurate?" How does one determine whether the decision was "fair?"

[7] The above questions illustrate the need for a principled review of algorithmic decision-making used by governmental actors to ensure that procedural due process safeguards are adequate and just when people are deprived of constitutionally-protected rights and interests. Part II of this Article provides background on the algorithmic decision-making process and the ways that AI technologies can undermine procedural due process protections. Part III explains the ways that AI technologies can enhance due process protections. Part IV explores other jurisdictions, such as the EU and UK, for potential solutions that are both efficient and fair – two characteristics that often conflict with each other. Finally, Part V provides a checklist of questions and issues to consider when challenging government decisions made by AI.

---

[30] *See* Bloch-Wehba, *supra* note 5, at 1274–75 (explaining that since state Medicaid agencies began privatizing their decision-making and relying on algorithms to decrease recipients' benefits, states have faced an influx of due process claims).

[31] 42 U.S.C. § 1396(a)(3); Bloch-Wehba, *supra* note 5, at 1275–76 (explaining that any suspension, reduction, or termination of benefits requires an opportunity for a hearing); *see also id.* at 1276 ("Congress also required state agencies to base Medicaid waiver budgets on a 'methodology that uses valid, reliable cost data [which] is open to public inspection, and includes a calculation of the expected cost of such services.'").

II. How Algorithmic decision-making Processes Undermine Procedural Due Process

A. A Brief Primer on Algorithmic Decision-Making Terminology

[8] There are four basic types of AI systems: descriptive, diagnostic, predictive, and prescriptive.[32] Descriptive systems rely upon "nominal" classifications, which make decisions based on identifying certain specified characteristics.[33] An example is a binary, yes/no, or 0/1 response. Diagnostic systems focus on quantification and sometimes use "cardinal" classifications.[34] Predictive systems are governed by ranking or ordinal classification systems, and they involve value judgments to determine relative strengths (risks and rewards) among those who are classified.[35] Prescriptive systems address the larger question of what "should be done," and involve more complicated layers of analytics, including regressions.[36] Determining which type(s) of system(s) to use to address a problem is one of the first steps in crafting algorithms. With each of these types of problems, AI can provide effective measurements and reliable assessments, but when AI is used for other types of problems, "it is quite difficult to establish their efficacy and reliability."[37]

---

[32] *See* Iria Giuffrida, *Liability for AI Decision-Making: Some Legal and Ethical Considerations*, 88 FORDHAM L. REV. 439, 440 (2019).

[33] Burk, *supra* note 24, at 1193.

[34] *Id.*

[35] *See id.*; Paul W. Grimm et al., *Artificial Intelligence as Evidence*, 19 NW. J. TECH. & INTELL. PROP. 9, 24 (2021).

[36] Grimm et al., *supra* note 35, at 25.

[37] *Id.* at 31 (expressing concern when artificial intelligence techniques are used to "cluster date, to detect anomalies, or to predict exceedingly rare events").

[9] There are three phases in the development of the AI decision-making process.[38] The first is the "goal-setting phase," the second is the "coding phase," and the third is the "implementation phase." During the goal-setting phase, designers and software engineers figure out what type of system to create, and what the goals of the system will be (for instance, to determine the amount of benefits a bedridden veteran is entitled to, based on the extent of his disability), how conflicting variables should be ranked in the hierarchy (if the disability is only partially service-related, for instance) and what features will be measured, which inputs will be gathered, and what datasets will be used. When designers of the algorithm do not understand the task at hand well enough, they might create an inadequate design.

[10] The coding phase provides the blueprint for the order of operations and decision matrices, and it also addresses the potential for bugs and biases, as well as accuracy rates and other performance metrics. In the implementation phase, actual decisions are made. In the case of machine learning,[39] these decisions are fed back into phases one and two for post-audit training of the algorithm, and subsequent retraining, to help the machine learn.[40] When machine learning algorithms are developing, they must "balance two objectives: *exploration*, in which it learns as much as it can, and *exploitation*, in which it employs what it has learned thus far to address the problem at hand."[41]

[38] Karl de Fine Licht & Jenny de Fine Licht, *Artificial Intelligence, Transparency, and Public Decision-Making*, 35 AI & Society 917, 919 (2020).

[39] *See* Giuffrida, *supra* note 32, at 441–42 ("[T]he basic steps in the creation of an AI based on machine learning are: (1) coding of the underlying AI program; (2) training the AI to accomplish its function; and (3) ongoing self-modification by the AI based on changes in the underlying data and feedback loops.").

[40] Licht & Licht, *supra*, note 38 at 920.

[41] Grimm et al., *supra* note 35, at 25–26.

[11] Coding is sometimes referred to as "playing with the data," whereas implementation is the "running model," and each phase presents different challenges.[42] Conflating the two can lead to erroneous conclusions. For instance, when some say, "we must preserve a 'human in the loop' of machine learning, [] most of them are referring to the running model as the relevant loop," but it may be as or more important "to maintain humans in the underappreciated playing-with-the-data loop as well."[43]

[12] Two of the main reasons why algorithms fail is because they do not fit the data, or they may be unable to generalize and perform in real-world situations. There may be cases that just do not fit into any of the classification models, and finally, there is "a class of cases in which there is no outcome variable available that is well enough correlated to the underlying variable of interest."[44] In these situations, AI is not well-suited, and humans need to be involved in making that determination.

[13] With this background in mind, we now turn to the Due Process Clause.

B. What Kind of Hearing Does AI Due Process Require?

[14] Debates over the meaning and nature of due process required to satisfy constitutional mandates abound.[45] In a 1975 *Pennsylvania Law Review* article, Judge Friendly began with a kernel of wisdom from a 1974 opinion by Justice White, arguing that "[t]he Court has consistently held that *some kind of hearing* is required at some time before a person is finally

---

[42] Lehr & Ohm, *supra* note 7, at 655.

[43] *Id.* at 657–58.

[44] Aziz Z. Huq, *Constitutional Rights in the Machine-Learning State*, 105 CORNELL L. REV. 1875, 1913 (2020) (testing against due process norms).

[45] *See, e.g.*, Michael J. Perry, *Constitutional "Fairness": Notes on Equal Protection and Due Process*, 63 VA. L. REV. 383, 419–20 (1977) (explaining Constitutional debates surrounding when due process rights vest in contemporary period of the 1970s).

deprived of his property interests."[46] After briefly addressing when a hearing is necessary, Judge Friendly focuses on determining the nature of a constitutional hearing. The elements of a fair hearing are: 1) an unbiased tribunal, 2) notice of the proposed action and the grounds asserted for it, 3) an opportunity to present reasons why a proposed action should not be taken, 4) the right to call witnesses, 5) the right to know evidence against oneself, 6) the right to have decisions based only on the evidence presented, 7) the right to counsel, 8) making of a record, 9) statements of reasons, 10) public attendance, and 11) judicial review.[47]

[15] Algorithmic decision-making and the use of machine learning technologies violate most—if not all—of these identified elements of a fair hearing. In a recent *Columbia Law Review* article, Margot Kaminski interprets Judge Friendly's elements as "more of a menu than a checklist; what constitutes a fair hearing may vary with circumstances such as the level of harm and the administrative costs."[48]

[16] While *Mathews v. Eldridge* held that a paperwork review process would be adequate process for a denial of Social Security benefits,[49] it established a framework based on three factors for determining whether the government had satisfied due process: (1) the private interest, (2) the risk of erroneous deprivation, and (3) the governmental interest (especially fiscal and administrative).[50] Given the time savings over holding individual pre-

---

[46] Friendly, *supra* note 1, at 1267 (emphasis added).

[47] *Id.* at 1279–82, 1287, 1291, 1293–94.

[48] Margot E. Kaminski & Jennifer M. Urban, *The Right to Contest AI*, 121 COLUM. L. REV. 1957, 2035 (2021) (arguing for an individual right to challenge AI decisions based on principles of due process but modified for the digital age and positing the right to contest AI decisions accords with due process theory).

[49] Mathews v. Eldridge, 424 U.S. 319, 348–49 (1976) (reasoning that a pretermination hearing before an administrative official was not required).

[50] *Id.* at 334–35.

deprivation hearings,[51] agencies must still demonstrate that algorithmic deprivations have substantially low error rates under the second factor.[52]

[17] Understandably, individual claimants want to be assured that the decision was correct in their own cases, without regard to overall error rates or the cases of others.[53] But the risk of erroneous deprivation is not enough; to confer standing, courts have found that "[t]he speculative risk that at some future point some individual *might* lose funding is not a basis for recovery and indeed does not even provide standing."[54] A mere assertion that an algorithm *might* result in unfair benefits is not sufficient to provide standing for a due process challenge.[55] While errors in algorithms can impact hundreds and even millions of potential claimants, as Professor Huq

---

[51] *See* Huq, *supra* note 44, at 1909–10 (explaining how the high costs associated with individual hearings create additional challenges).

[52] *See* Cary Coglianese & David Lehr, *Regulating by Robot: Administrative Decision Making in the Machine-Learning Era*, 105 GEO. L.J. 1147, 1186 (2017).

[53] *See* Huq, *supra* note 44, at 1909 (providing that the difficulty in safeguarding due process interests, however, is that litigants generally are disinterested in uncovering systemic problems with algorithmic instruments; rather, their interests lie in revealing the errors in their individual cases).

[54] Brandy C. v. Palmer, No. 17cv226, 2018 WL 4689464, at *2 (N.D. Fla. Sept. 29, 2018).

[55] *Id.* at *4. This case evaluated algorithmic due process in the context of disability benefits. *See id.* at *2. The plaintiffs obtained benefits from Florida through the state's program, Agency for Persons with Disabilities, and argued that the Agency put them in danger of institutionalization when it relied on the "iBudget" automated system for assessing recipients' benefits. *Id.* at *1–2. The court evaluated whether the algorithm was "fatally flawed" based on the standard set forth in *Olmstead v. L.C. ex rel. Zimring*, 527 U.S. 581 (1999) (holding that a state violates the Americans with Disabilities Act if it unnecessarily makes isolation in an institution a condition of public assistance for individuals with disabilities). *Id.* at *1, *3. The court, reasoning that the ultimate concern "is not how a state arrives at a benefit amount but whether the state provides an adequate benefit," ruled that there was no evidence that any individual had lost assistance or was likely to lose assistance because of the algorithm's methodology. *Id.* at *3–4.

asserts, the "individualized hearing model, however, is not well suited to the identification of such systemic problems," and once the flaws are found, a hearing may not even be necessary, as the error may only become evident in the aggregate.[56] Moreover, the program can be adjusted and the input rerun to reach the proper outcome.

[18] Professor Huq explains that many have read *Mathews* "to demand that specific notice be given to regulated subjects and that an individualized determination, often involving a human adjudicator, be available,"[57] which may "miss the best way to vindicate due process interests for a number of reasons."[58] Perhaps a better approach is that suggested by John Villasenor and Virginia Foggo, who derive three principles from the *Mathews* factors as applied to AI: auditability, transparency, and consistency.[59] Professor Kaminski apparently agrees, stating that a "right to contest AI that does not include at least elements of notice, evidentiary disclosure, and reason-giving will not provide a meaningful hearing."[60]

[19] The next subsection discusses these principles in relation to the various components of what constitutes a "fair hearing."

---

[56] *See id.* at 1909–10.

[57] *Id.* at 1908.

[58] *Id.* at 1909.

[59] John Villasenor & Virginia Foggo, *Artificial Intelligence, Due Process, and Criminal Sentencing*, 2020 MICH. ST. L. REV. 295, 351–52 (2020) (analyzing these three principles' additional burdens on governments and on private companies that supply risk-assessment software to government agencies).

[60] Kaminski & Urban, *supra* note 48, at 2035–36 ("[I]ndividuals are unlikely to feel respected by a contestation right that does not provide a sufficient window into decision-making—through notice, evidence, and reason giving—to make meaningful challenges possible.").

C. The Elements of a Fair Hearing for Algorithmic Decision-Making Outcomes

[20] The general components of a fair hearing have been examined in great detail over the past century, and now, scholarship is emerging on the implications of AI technologies on fairness. This section provides a deeper dive into what it means to provide a fair hearing on algorithmic decisions that result in deprivations.

1. An Unbiased Tribunal

[21] Despite the best efforts of programmers and software engineers, an algorithm may be trained on incomplete, biased, or flawed data, or there may be data black holes (due to a lack of enforcement perhaps) and thus no reported cases or data to train upon for certain groups or circumstances.

[22] There are several types of biases implicated in machine learning processes.[61] *Historical bias* arises when the world, as it is, is biased (e.g., men-dominated image search results for the word "CEO" simply reflect that 95% of Fortune 500 CEOs actually are men).[62] *Representation bias* occurs when some groups of the population are underrepresented in the training dataset.[63] For example, models trained on ImageNet, where 45% of images come from the United States and only 1% of images represent China, perform poorly on images depicting Asians.[64]

---

[61] *See* Harini Suresh & John V. Guttag, *A Framework for Understanding Unintended Consequences of Machine Learning*, EQUITY & ACCESS ALGORITHMS, MECHANISMS, & OPTIMIZATION, Oct. 2021 (explaining that there are five sources of bias in machine learning).

[62] *See id.*

[63] *Id.*

[64] *Id.*

[23] *Measurement bias* arises when there are issues with choosing or measuring the particular features of interest.[65] The issues may come from varying granularity or quality of data across groups or oversimplification of the classification task.[66] For example, a GPA score is often used as a measure of student success, though it does not capture many important indicators of success.

[24] *Aggregation bias* occurs "when a one-size-fit-all model is used for groups with different conditional distributions." For example, studies suggest that HbA1c levels, which are used for diagnosing diabetes, differ in a complex way across ethnicities and genders.[67] Thus, a single model is not likely to be the best fit for predicting diabetes for every group in the population. *Evaluation bias* arises when evaluation or benchmark datasets are not representative of the target population.[68] Unrepresentative datasets encourage the development of models that only perform well on a subset of data. For example, facial recognition benchmarks used to have a very small fraction of images with dark-skinned female faces, which resulted in commercial facial recognition systems performing very badly on this subset of the population.[69]

[25] And finally, *deployment bias* occurs after model deployment, when a system is used or interpreted in inappropriate ways.[70] Deployment bias is often a concern when systems are used as decision aids for humans because the human intermediary may act on AI-generated predictions in ways that

[65] *See id.*

[66] Suresh & Guttag, *supra* note 61.

[67] *Id.*

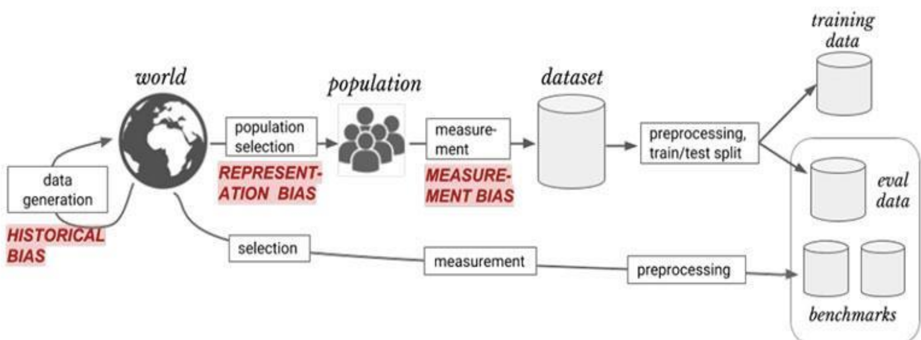
[68] *Id.*

[69] *Id.*

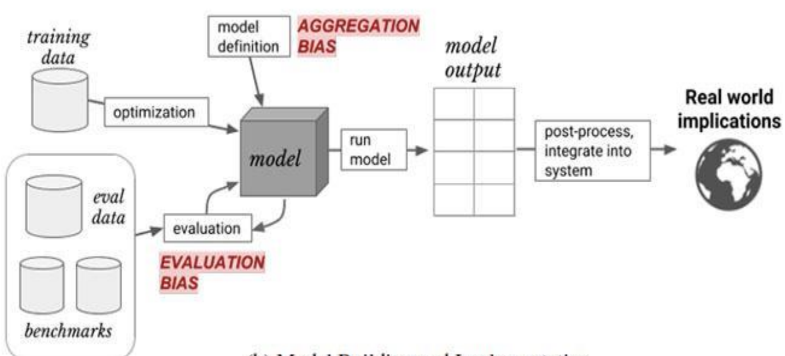
[70] *See id.*

are typically not modeled in the system.[71] For example, risk assessment tools in the criminal justice system predict a risk score, but a judge may interpret this in unexpected ways before making his or her final decision, thus leading to a result that conflicts with the machine-generated recommendation.[72]

[26] The Figure below, reproduced from Harini Suresh and John Guttag's article, provides a graphic representation of where biases manifest in machine learning processes.[73]



(a) Data Generation



(b) Model Building and Implementation

[71] See Suresh & Guttag, *supra* note 61.

[72] *Id.*

[73] *Id.*

2. Notice of the Proposed Action and the Grounds Asserted for It

a. Reasons for the Decision

[27] Danielle Citron's early work identified some of the specific dangers to due process that arise through automated systems, particularly those stemming from the blurred line between rulemaking and adjudication.[74] The basis for the decision is often obscured because as the machine makes decisions, it adjusts its 'rules' to make 'better' decisions in subsequent cases. The reasons must logically be connected to the rules as applied, not just as written. Algorithmic decision-making presents special challenges to agency decision-making systems,[75] in part due to the lack of transparency and the failure of system developers to provide audit trails,[76] two of the principles articulated by Villasenor and Foggo.[77] Agencies using algorithms rarely can provide details about the reasons why a deprivation was ordered,[78] due to the process' lack of transparency and the often-proprietary nature of the algorithms themselves. While the decision—the outcome—is clear, the rationale supporting it often is not.

[74] Citron, supra note 27, at 1278–81 (explaining the Colorado public benefit system and how "factual errors and illicit rules," those that play a role in an algorithm's assessment even though the policymakers did not make that requirement, lead to denial of benefits).

[75] *See* Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1, 28 (2014).

[76] *Id.* at 24, 28, 33 ("Technological due process insists that automated systems include immutable audit trails to ensure that individuals receive notice of the basis of decisions against them."); *see also* Coglianese & Lehr, *supra* note 53, at 1205–07 (explaining that legitimacy requires administrative agencies to provide adequate reasons for their actions, assumptions, and methodologies).

[77] Villasenor & Foggo, *supra* note 59, at 339, 343, 347.

[78] *See* Citron & Pasquale, *supra* note 75, at 10–11.

[28] For instance, in *Michael T. v. Bowling*, West Virginia declined to provide information about the inputs, variables, and weight of factors used in denying Medicaid benefits.[79] The court held that the algorithm's lack of transparency posed an unreasonable risk of erroneous deprivation and foreclosed the agency from its further use.[80] Similarly, when an agency declines to provide reasons for denying an informal review of its decisions, courts find due process violations.[81]

[29] A "missing algorithm" jury instruction like the one that Deborah Won proposes in criminal cases, which is similar to the "missing witness" instruction, would help to balance a person's due process rights with intellectual property protections.[82] Her proposed instruction would help to remedy *Brady* conflicts; a similar policy or presumption could apply in administrative proceedings to provide a presumption, or burden-shifting, where the agency was unwilling or unable to persuade the private vendor to disclose the algorithmic basis for the denial of benefits.[83] One benefit of this

---

[79] Michael T. v. Bowling, No. 15-cv-09655, 2016 U.S. Dist. LEXIS 123749, at *6–7, *10 (S.D. W. Va. Sep. 13, 2016) (alluding to a claim predicated upon West Virginia's Home and Community-Based Services (HBCS), under which the State did not publicize the algorithm used in budget allocation determination, despite due process mandating they do so).

[80] *See id.* at *42; Bloch-Wehba, *supra* note 5, at 1278.

[81] *See, e.g.*, K.W. v. Armstrong, 180 F. Supp. 3d 703, 715 (D. Idaho 2016) (reasoning that due process required IDHW to include its reasons for denying informal review and to contact an appropriate representative, like a family member or guardian, before proceeding with an appeal, but since none of these measures were taken, IDHW failed to meet the due process requirements).

[82] Deborah Won, *The Missing Algorithm: Safeguarding* Brady *Against the Rise of Trade Secrecy in Policing*, 120 MICH. L. REV. 157, 186 (2021) (describing the application of algorithms in law enforcement and the relationship between trade secrets and artificial intelligence and arguing that trade secrets conflict with the Due Process Clause).

[83] *Id.*

approach is that it avoids "a solution as extraordinary as excluding algorithmic evidence altogether."[84]

b. Inadequate Notice

[30] The lack of clear and effective notice for algorithmic decision-making is another threat to due process. While some systems do not provide any pre-deprivation notice,[85] for those that do, the lack of an audit trail often leads to inadequate notice in terms of a warning prior to the expected loss of benefits or other adverse decisions. An Arkansas court expanded on the extent of notice required for algorithmic decisions in *Jacobs v. Gillespie*.[86] The court held that due process requires notice which is "as specific as reasonably practicable" when providing a rationale for a benefit reduction, "with specific references (as applicable) to the beneficiary's [algorithmic] assessment, the beneficiary's [program], and the [automated system], including the algorithm."[87] But would providing the algorithm be of any assistance to the average welfare recipient? Likely not, given the complexities involved. Even though the beneficiaries' assessment may include a numeric score, without knowledge of the range, mean, and distribution of scores within the beneficiary population, the reasons for the score may remain opaque.

[84] *Id.*

[85] *See* Bloch-Wehba, *supra* note 5, at 1278.

[86] *See* Estate of Jacobs v. Gillespie, No. 16-cv-119, 2017 U.S. Dist. LEXIS 106754, at *2–3 (E.D. Ark. July 11, 2017).

[87] Order, 2017 U.S. Dist. LEXIS 106754 (2017) (ordering that notice regarding algorithmic reductions in attendant care hours must provide ascertainable standards only regarding the content of information to demonstrate the state's decision-making process).

3. An Opportunity to Present Reasons Why a Proposed Action Should Not Be Taken

[31] Participation is essential to people's notions of fundamental fairness because it provides them with the opportunity to explain.[88] Protestations about the risk of harm can be overlooked when humans are not in the loop of algorithmic decision-making.[89] In *K.W. v. Armstrong*, disabled plaintiffs argued that the algorithmic system's outcomes would make such severe cuts to their benefits that they would be at "serious risk of institutionalization."[90] The lawsuit was necessary for their claims to receive the hearing that due process requires.[91]

4. The Right to Call Witnesses

[32] The right to call witnesses to explain is also paramount in the fairness analysis.[92] Witnesses must testify from personal knowledge or provide lay or expert opinions.[93] Even when a decisional error is clearly identified, agency officials are unable to testify from personal knowledge because they cannot identify the processing problems—let alone understand how they arose—and therefore have no basis for testimony. In contrast, under French law, public officials who use algorithms must "be able to exercise control over them and to explain how they work to an affected

[88] Kaminski & Urban, *supra* note 48, at 2037 ("It will depend on whether the contestation process and timeline are clear and low cost").

[89] *Id.* at 1973.

[90] K.W. v. Armstrong, 180 F. Supp. 3d 703, 722 (D. Idaho 2016).

[91] *Id.* at 715 (explaining that under the *Mathew*s balancing test, due process required the Idaho Department of Health and Welfare to provide specific written standards and regulation that defined "health and safety").

[92] *See* Kaminski & Urban, *supra* note 48, at 2037.

[93] FED. R. EVID. 602, 701–02.

person."[94] Without such a requirement, when conditions change, as Calo and Citron note, "the official is not in a position to adapt," thus forestalling any use of discretion as delegated by Congress.[95] They propose "the deliberate and self-conscious adoption of technology *to the extent* it furthers the rationales for delegating authority and power to agencies and not otherwise."[96] How does one determine the extent to which specific AI technologies enhance agency legitimacy? That is the question that Part III addresses.

5. The Right to Know Evidence Against One

[33] On this decision-making aspect, a figure entitled *Notional Rendition of Decision Pipeline for the U.S. Criminal Justice System*, provides an interesting flowchart of opportunities for decisions to be made.[97] The authors provide some useful examples of violations of due process in the knowledge of the precipitating circumstances context of benefits being curtailed, such as the cases involving the "No-Fly List."[98] In a similar case recently addressed by the U.S. Supreme Court, individuals whose credit reports erroneously identified them as persons with which it was illegal to do business were denied access to the information that formed the bases for including the erroneous information in the reports when they applied for credit.[99] This unwarned deprivation exacerbates automation bias, which leads individuals to be less likely to overturn a decision once a machine has

---

[94] Kaminski & Urban, *supra* note 48, at 2038.

[95] Calo & Citron, *supra* note 13, at 832.

[96] *Id.* at 837–38.

[97] OSONDE A. OSOBA ET AL., ALGORITHMIC EQUITY: A FRAMEWORK FOR SOCIAL APPLICATIONS 45 (2019).

[98] *Id.* at 48.

[99] TransUnion LLC v. Ramirez, 141 S. Ct. 2190, 2214 (2021).

made it.[100] Heeding the machine-generated warning, and without further investigation, the car dealers and the department store credit card issuers deprived the plaintiffs of fair access to credit.

6. The Right to Have Decisions Based Only on the Evidence Presented

[34] On the issue of evidence presented, Danielle Citron notes that even a hearing in front of an administrative law judge or other official may be inadequate because automation bias may influence the hearing officer's willingness to exercise discretion in the face of a contrary algorithmic decision or recommendation.[101] On the other hand, some people hold algorithms to higher standards, and thus systematically dismiss them in the face of any error.[102]

[35] Applying the *Mathews* balancing test, Citron notes that courts would be unlikely to find in favor of a claimant seeking access to the automated program source code—for instance, by providing for a sufficiently qualified expert to interpret that code for the claimant. She argues that the *Mathews* calculation may be inappropriate in this context because once the system is fixed, there is no need for further expert witness costs because automation bias is not a "one-off" as is the case for human decision-making in benefit calculations.[103] The missing source code (analogous to the missing witness instruction notes above) could be an appropriate redress to negate the impact of automation bias.

---

[100] OSOBA ET AL., *supra* note 97, at 48.

[101] *See* Citron, *supra* note 27, at 1283 (explaining that even though a hearing officer is a safeguard for due process, hearing officers are still influenced by automation bias).

[102] *See* Grimm et al., *supra* note 35, at 47.

[103] *See* Citron, *supra* note 27, at 1284–85.

7. The Right to Counsel

[36] Only a few brief words are necessary here. Who can provide legal advice on how best to position oneself in an algorithmic decision-making process? The software engineer who designed and trained the AI likely can provide better guidance than an attorney who does not understand the technology. This question will be the subject of another article.

8. The Making of a Record

[37] Numerous sources have addressed the need for transparency and audit trials to preserve a record of the decision-making process.[104]

9. Statements of Reasons

[38] Another danger is inadvertent rulemaking.[105] Rules, as applied, are more salient in the AI context than rules as written. Policies become rules if they are followed in the vast majority of situations. Thus, an "automated system's application of distorted policy in hundreds of thousands of cases similarly can be seen as establishing a new rule."[106] The 'reason' for the application of the distorted policy becomes a latent factor that the algorithm finds useful and efficient, but the impacted person may never uncover this 'deciding factor.' Eventually this factor becomes a *de facto* 'element' and

---

[104] *See* Andrew Burt, *The AI Transparency Paradox*, HARV. BUS. REV. (Dec. 13, 2019), https://hbr.org/2019/12/the-ai-transparency-paradox [https://perma.cc/ZR76-BJMP] (noting both the need for transparency and the risks associated with it); Sylvia Lu, *Algorithmic Opacity, Private Accountability, and Corporate Social Disclosure in the Age of Artificial Intelligence*, 23 VAND. J. ENT. & TECH. L. 99, 158 (2021) ("[I]t is crucial to ensure that algorithms are harmless by requiring more transparency to mitigate algorithmic opacity"); Peter K. Yu, *Beyond Transparency and Accountability: Three Additional Features Algorithm Designers Should Build into Intelligent Platforms*, 13 NE. L. REV. 263, 274 (2021).

[105] *See* Citron, *supra* note 27, at 1288.

[106] *Id.* at 1289.

therefore part of the rule. As agency rules are subject to notice and comment,[107] algorithmically-created rules should be subject to notice and comment, and algorithmic decisions that result from machine learning violate this mandate[108] and can also compromise the principle of "consistency" articulated by Villasenor and Foggo.[109]

10. Public Attendance

[39] Public attendance at a hearing is one component of due process that is difficult to balance with the use of symbolic systems, but a failure to provide an opportunity for public discourse on decision making can also violate due process.[110]

[107] *See* Josh Armstrong, *Necessary "Procedures": Making Sense of the Medicare Act's Notice-and-Comment Requirement*, 87 U. CHI. L. REV. 2175, 2179–80 (2020) ("[A]n agency seeking to adopt a new administrative rule ordinarily must first give notice of the proposed rule and subject it to a period of public comment . . . The APA makes exceptions to its notice-and-comment requirement, however, for 'interpretative rules, general statements of policy, or rules of agency organization, procedure, or practice").

[108] *See* OSOBA ET AL., *supra* note 97, at 11; Adam Finlay & Catherine Walsh, *Council Publishes Proposed Amendments to Draft AI Regulation*, MCCANN FITZGERALD (Dec. 23, 2021), https://www.mccannfitzgerald.com/knowledge/technology-and-innovation/council-publishes-proposed-amendments-to-draft-ai-regulation [https://perma.cc/2SAP-U8SF] (including specific publications on using AI systems that exploit socially or economically vulnerable people or groups); Citron, *supra* note 27, at 1290–93 ("Unfortunately, the opacity of code makes it difficult to determine if a change has imposed a 'new rule' requiring rulemaking procedures or an 'interpretive rule' arguably demanding less process.").

[109] *See* Villasenor & Foggo, *supra* note 59, at 347.

[110] *See* Ark. Dep't of Hum. Servs. v. Ledgerwood, 2017 Ark. 308, at 13–14, 530 S.W.3d 336, 345 (discussing a group of plaintiffs who opposed the state's passage of a rule changing the process used to decide attendant-care hours for Home-and Community-Based Services (HCBS) participants).

11. Judicial Review

[40] How much do courts understand about AI systems? Most courts understand very little; however, this topic is explored in depth in other scholarship.[111]

III. WAYS AUTOMATED SYSTEMS CAN ENHANCE DUE PROCESS

[41] There are several ways to promote procedural due process protections in government uses of AI technologies. First, through efficient monitoring of processes; second, with transparency; and third, by focusing on an anti-stereotyping goal. Remember, the status quo is not neutral.[112]

A. Efficient Monitoring

[42] A utilitarian perspective, which highlights that the "right action is the one that maximizes overall utility . . . provides a straightforward fairness-assessment calculus, depending on whether or not the use of the algorithm maximizes utility."[113] "For a utilitarian, decisionmaking processes that use algorithms matter only as they relate to valued outcomes,"[114] and those that promote efficiency can maximize utility.

---

[111] *See* Ashley Deeks, *The Judicial Demand for Explainable Artificial Intelligence*, 119 COLUM. L. REV. 1829, 1838 ("Judges are well positioned in this ecosystem to develop pragmatic approaches to xAI, even though they are not—indeed, because they are not—experts in machine learning technology.").

[112] *See* Wachter, et al., *supra* note 4, at 767 (introducing the concept of the status quo's non-neutrality within the context of machine learning).

[113] OSOBA ET AL., *supra* note 97, at 11.

[114] *See id.* at 11–15 (discussing statistical conceptions of equity, including the notions of fairness through unawareness, individual fairness, demographic parity, and asymmetry in decision-making costs).

[43] On the efficiency monitoring point, Danielle Citron makes the case for technological due process, and describes ways to re-conceptualize procedures to: 1) secure meaningful notice, 2) provide protections for hearings, and 3) maintain proper rulemaking procedures.[115] First, requiring audit trails would mean that individuals could obtain the rationale for the automated decision.[116] The second point requires training hearing officers on how to recognize and reduce the impact of automation bias,[117] and how to reconceive the *Mathews* test in situations where "retrofitting an automated system's reasoning was essential to enabling individuals to address an agency's intended actions."[118] Third, requiring open source code[119] means that any improper, implicit rulemaking can be discovered, redressed, and redacted, with substantial trial run testing.[120]

B. Transparency

[44] Enhancing transparency can promote due process.[121] "Transparency" can be achieved when "government (A) can make its source code available (B) to the public (C) so that they can see that nothing untoward is occurring in relation to their use of an algorithm to better [make

[115] *See* Citron, *supra* note 27, at 1305–13.

[116] *Id.* at 1305.

[117] *Id.* at 1306.

[118] *Id.* at 1308.

[119] *See id.* at 1308–10.

[120] *See* Citron, *supra* note 27, at 1310–11.

[121] *See* Licht & Licht, *supra* note 38, at 918; *see also id.* ("Based on our reading of the literature, we argue that a limited form of transparency that focuses on *providing justifications* for decisions has the potential to provide *sufficient grounds* for perceived legitimacy in AI decision-making.").

predictions].”[122] There are three types of transparency: one that informs about final decisions or policies, a second that informs about the process resulting in the decisions, and a third that informs about the reasons.[123] Employing only one of the three types may not lead to a perception of greater legitimacy for the use of artificial intelligence, so the public bodies must determine what type of transparency they seek to promote.[124]

[45] More transparency in the process is not necessarily better in all circumstances; in fact, transparency can lead to gaming the system by those who understand how to manipulate the variables. Providing the public the information it needs means giving it information that it can understand, and full transparency of the process may not be understandable to those without the relevant education. Conversely, if the process is too transparent, some members of the public may be skeptical of the simplicity and think that the transparency is actually obfuscation. Efficiently informing the public while considering these three aspects of transparency remains a challenge.[125]

C. Anti-Stereotyping

[46] An anti-stereotyping approach could lead to greater procedural protections for individuals, and it may satisfy the public’s idea of fairness. With algorithms, individuals are judged in comparison to group data—this

[122] Licht & Licht, *supra* note 38, at 918.

[123] *Id.*

[124] *See id.* at 919 (“[*I]f* perceived legitimacy is the goal, we should opt for transparency in rationale and not transparency in process. . . . [I]f explicability means that we actually make the decision-making processes fully transparent, then we do not believe it suitable in relation to the production of perceived legitimacy[.]”).

[125] *See id.* at 924.

is the essence of stereotyping.[126] Bornstein sets forth an anti-stereotyping theory which requires that one not only be treated equally, but also individually, under the law.[127] The notion of individual fairness means that similar individuals obtain similar outcomes.[128]

[47] On the issue of fairness through unawareness, which is a way of simulating Rawls' veil of ignorance, one could argue that the decision-making process must be neutral "with respect to specified individual sensitive attributes. These attributes often include race, gender, or other attributes" but sometimes exclude socioeconomic status and cultural factors.[129] One critique, however, is that it is important for AI to recognize and consider some of these sensitive attributes to improve the effectiveness and efficiency of the algorithms.[130] The model can then learn to be fairer to individuals if their attributes are noted as well.[131]

[48] Definitions for fairness are more elusive in symbolic systems. Demographic parity relies upon mirroring distributions within an overall population aiming to equitably distribute resources. Equal opportunity can equalize "accuracy rates and misclassification rates across all classes."[132] Counterfactual fairness means that the "outcome remains unchanged when

[126] *See* Stephanie Bornstein, *Antidiscriminatory Algorithms*, 70 ALA. L. REV. 519, 526 (2018) ("The fact that a computer, instead of a human, does the stereotyping should not insulate from liability the employer who relies on the stereotyped results if the employer's intentional use of an algorithm discriminates.").

[127] *See id.* at 544.

[128] *See* OSOBA ET AL., *supra* note 97, at 13.

[129] *Id.*

[130] *See id.* at 14.

[131] *See id.*

[132] *Id.*

sensitive attributes are changed in the causal world model."[133] Which definition we design the algorithm to maximize is what in effect becomes the 'rule.' Each aspect results in a different classification for what constitutes a 'fair' or 'just' outcome.

[49] An anti-stereotyping lens would alleviate many of the procedural problems addressed in Part II. In the employment context, Professor Bornstein suggests that existing Title VII doctrine, if focused on its anti-stereotyping goal, could make progress towards creating "actively *anti*discriminatory algorithms."[134] She catalogues some of the existing literature on algorithmic discrimination,[135] and notes that most of the approaches fall into the "improve the algorithms" or "improve the law" schools of thought, in part because they are approaching the problem from either a formal equality or substantive equality framework, instead of considering the anti-stereotyping approach.[136]

[50] The anti-stereotyping approach also means that individuals should not be judged based on stereotypes associated with any protected classes to which they belong.[137] The late Justice Ruth Bader Ginsburg used this

[133] OSOBA ET AL., *supra* note 97, at 15.

[134] Bornstein, *supra* note 126, at 520, 526 ("Applying an antistereotyping lens to the issue of algorithmic decision-making calls into question the underlying 'neutrality' of algorithms and the big data on which they rely.").

[135] *Id.* at 533–40; *see, e.g.*, Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CAL. L. REV. 671 (2016); Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. PA. L. REV. 633 (2017); Pauline T. Kim, *Data-Driven Discrimination at Work*, 58 WM. & MARY L. REV. 857 (2017).

[136] *See* Bornstein, *supra* note 126, at 520.

[137] *See id.* at 545 ("The [Supreme] Court has interpreted Title VII to protect an individual against protected class discrimination even when other members of the protected class, or the protected class as a whole, may not have suffered harm.").

approach in the gender discrimination cases she argued prior to joining the bench.[138]

[51] Professor Bornstein suggests that applying an anti-stereotyping lens could prove beneficial in one additional way: "it may help sort stereotype-activating uses of algorithms from stereotype-suppressing uses . . . "[139] This practice is also referred to as bias-preserving, versus bias-transforming metrics.[140] Such uses depend on the design and output goals of the algorithm when created.[141] She uses the example of an algorithm designed to determine whether someone is likely to move within the next ninety days.[142] Women—particularly women with children—may be most disadvantaged if that is the basis for whom an employer chooses to interview.[143] While the criteria may be job related, algorithms may not fulfill the business necessity prong of the test from *McDonnell Douglas Corp. v. Green*.[144] One final

[138] *See id.* at 545–46 ("[A] plaintiff may make out a case of disparate treatment under stereotype theory without comparator evidence because, where a work-related decision is made on the basis of a stereotype associated with a protected class, that 'can by itself and without more be evidence of an impermissible, [protected class]-based motive.'"). *See generally* Price Waterhouse v. Hopkins, 490 U.S. 228 (1989) (giving a quick review of a major gender discrimination case).

[139] Bornstein, *supra* note 126, at 551.

[140] *See* Wachter, *supra* note 4, at 42.

[141] *See id.* at 28 (using group fairness, conditional demographic parity, and counterfactual fairness as examples).

[142] *See* Bornstein, *supra* note 126, at 551.

[143] *See* McDonnell Douglas Corp. v. Green, 411 U.S. 792, 800–01 (1973) (finding that the purpose of the language of Title VII is to assure equality in employment and eliminate discriminatory practices); Bornstein, *supra* note 126, at 551–52 ("Entelo advertises that its predictive matching algorithm allows employers to highlight diverse candidates to increase workforce racial or gender diversity.").

[144] Bornstein, *supra* note 126, at 555–56 (recommending that validation studies should perhaps cover content, rather than "criterion" or "construct").

criticism "current scholarship fails to recognize" is that an employer intentionally applying a system to individuals that starts from a "garbage in" position infected by protected class bias may constitute disparate treatment.[145] In particular, the use of predictive matching by algorithms to find applicants that fit a model "good employee" may pose a problem of protected class stereotyping.[146] If an employer creates a model employee based on past subjective decision-making that incorporates protected class stereotypes and then applies that model to each future applicant, seeking to hold each individual to the stereotype of "good employee," that may no longer be a "facially neutral" practice.[147] The fact that a computer is making the decisions instead of a human does not wash away prior bias or make its application "neutral."[148]

[52] Recognizing the impact of AI technologies on private employer hiring decisions, the city of New York recently enacted a regulation that addresses the unbiased tribunal and notice components of fair hearings.[149] While these hearings are not necessarily governmental decisions, (although they could be for government employers within the city of New York) this regulation will be an interesting foray into the laboratory of democracy and may become the beginning of a wave of additional regulations. The New York City regulation will require all employers within that city to refrain from using automated employment technology to screen job candidates unless or until that technology has been subjected to a bias audit conducted

---

[145] *Id.* at 562.

[146] *Id.*

[147] *Id.* at 563.

[148] *Id.* at 562–63; *see also id*. at 565 (concluding if an employer raises the algorithm as a "legitimate nondiscriminatory reason," then proving the algorithm is biased will be enough to establish pretext).

[149] Nicol Turner Lee & Samantha Lai, *Why New York City is cracking down on AI in hiring*, BROOKINGS (Dec. 20, 2021), https://www.brookings.edu/blog/techtank/2021/12/20/why-new-york-city-is-cracking-down-on-ai-in-hiring [https://perma.cc/2XS6-S3SX].

by an independent auditor.[150] Set to take effect in January 2023, the law would require employers to inform applicants if the technology was used in making a decision about their applications, as well as to provide notice of the expected use of these technologies so that candidates may opt for an alternative selection process or some other sort of accommodation.[151]

[53] This Title VII analysis provides a theory for addressing the right to an unbiased tribunal, and the right to have decisions based only on the evidence presented—not generalizations and stereotypes—as addressed in Part II.C.1 and 6. In a similar manner as with public benefits and other resource allocation decisions, reducing the use of algorithms that rely upon human stereotypes would promote individual fairness.

IV. GUIDANCE FROM OTHER JURISDICTIONS

[54] The EU and the European Commission suggest "four key principles underpinning the development of AI systems: respect for human autonomy, prevention of harm, fairness, and explicability."[152] These principles are crucial: First, "[i]ndeed, under European Union law, automated decisions that have legal or similar effects on individuals, as AI decisions may, are required to be subject to some type of human review."[153] Second, the EU proposal would require prior conformity assessment for uses of AI with

---

[150] Erin Mulvaney, *Artificial Intelligence Hiring Bias Spurs Scrutiny and New Regs*, BLOOMBERG L. (Dec. 29, 2021, 5:30 AM), https://news.bloomberglaw.com/daily-labor-report/artificial-intelligence-hiring-bias-spurs-scrutiny-and-new-regs [https://perma.cc/C8MC-8SK8].

[151] *Id.*

[152] Giuffrida, *supra* note 32, at 454–55.

[153] *Id*. at 445 n.38 (citing Regulation 2016/679, art. 22, 2016 O.J. (L 119) 1, 46 (EU)).

high risks of harm, and a sliding scale for those with medium or low risks.[154] Third, on the fairness issue, the recent modifications to the proposed rule define public consultation.[155] Fourth, medium risk technologies must comply with transparency requirements.[156]

[55] In determining how to allocate risk and assess liability for AI systems and negative outcomes, some significant problems manifest. "Theories of liability, along with legal and equitable remedies, are often founded on a certain belief about human motivation—the innate desire to avoid punishment. These remedies are arguably ill-suited for AI systems, and thus do not carry the same weight in shaping AI's behavior (assuming behavior is even the right term)."[157] Thus, a balancing test focusing on remedies relative to harm may be more appropriate.

A. The EU's Dec. 2021 Proposed Amendments to AI Regulation

[56] On April 21, 2021, the European Commission proposed an AI legal

[154] *See* David Matthews, *A new type of powerful artificial intelligence could make EU's new law obsolete*, SCI. BUS. (Dec. 21, 2021), https://sciencebusiness.net/news/new-type-powerful-artificial-intelligence-could-make-eus-new-law-obsolete [https://perma.cc/J9Y3-W39H]; Luca Bertuzzi, *EU Council presidency pitches significant changes to AI Act proposal*, EURACTIV (Dec. 2, 2021), https://www.euractiv.com/section/digital/news/eu-council-presidency-pitches-significant-changes-to-ai-act-proposal [https://perma.cc/PP8C-S828].

[155] Finlay & Walsh, *supra* note 108 (20 Nov 21).

[156] *See* Heike Felzmann et. al, *Transparency you can trust: Transparency requirements for artificial intelligence between legal norms and contextual concerns*, 6 BIG DATA & SOC'Y 1, 2–3 (2019), https://journals.sagepub.com/doi/pdf/10.1177/2053951719860542 [https://perma.cc/Q737-ZPXF].

[157] Giuffrida, *supra* note 32, at 451.

framework.[158] It requires AI systems that "pose an increased level of risk according to the commission's criteria [to] be subject to legal requirements."[159] The proposal differentiates among uses of AI, providing different guidelines based on whether a use has an "unacceptable, high, or low risk to human safety and fundamental rights."[160] The proposal was submitted for public comment and feedback.[161] On November 29, some significant amendments were proposed based on the commentary.[162] The amendments, once adopted, take effect twenty days after publication in the Official Journal and apply two years afterwards.[163]

---

[158] *See* Mark MacCarthy & Kenneth Propp, *Machines learn that Brussels writes the rules: The EU's new AI regulation*, BROOKINGS (May 4, 2021), https://www.brookings.edu/blog/techtank/2021/05/04/machines-learn-that-brussels-writes-the-rules-the-eus-new-ai-regulation [https://perma.cc/Q7U7-H74Y] (acknowledging the risks of AI systems).

[159] Eve Gaumond, *Artificial Intelligence Act: What Is the European Approach for AI?*, LAWFARE (June 4, 2021, 11:50 AM), https://www.lawfareblog.com/artificial-intelligence-act-what-european-approach-ai [https://perma.cc/8TD5-RTTB].

[160]Ana Hadnes Bruder et al., *The European Union Proposes New Legal Framework for Artificial Intelligence*, MAYER BROWN (May 5, 2021), https://www.mayerbrown.com/en/perspectives-events/blogs/2021/05/the-european-union-proposes-new-legal-framework-for-artificial-intelligence [https://perma.cc/9TMM-9MAP]; *see also* Charter of Fundamental Rights of the European Union, art. 34, 2012 O.J. (C 326) 391, 402 (establishing social security and social assistance as a fundamental human right).

[161] Finlay & Walsh, *supra* note 108.

[162] *Id.*

[163] Jetty Tielemans, *A look at what's in the EU's newly proposed regulation on AI*, IAPP (Apr. 21, 2021), https://iapp.org/news/a/a-look-at-whats-in-the-eus-newly-proposed-regulation-on-ai [https://perma.cc/6XU4-MHPJ]; *see also* Gaumond, *supra* note 159 (technologies grouped under "high-risk" include systems "embedded" in products under third-party control and "systems that are not embedded in other products"); Jay Modrall, *EU proposes new Artificial Intelligence Regulation*, NORTON ROSE FULBRIGHT (Apr. 2021), https://www.nortonrosefulbright.com/en/knowledge/publications/fdfc4c27/eu-to-propose-new-artificial-intelligence-regulation [https://perma.cc/8AAT-LM6E] (explaining "high-risk AI systems" include systems provide "access to and enjoyment of essential private services and public services and benefits").

[57] After the comments, the Council presidency proposed expanding the scope of AI regulation to include importers and distributors, product manufacturers who deploy the AI systems and authorized representatives.[164] The compromise proposal also created some additional exceptions to AI regulations, for instance, exceptions for AI used solely for scientific research and development, and AI use that does not result in anything being placed on the market or deployed into service.[165] In addition, this compromise text imposes upon both private and public actors a "prohibition [against] using AI systems for social scoring."[166] Further, it expands the definition of "vulnerable group" from racial and ethnic minorities to include "persons who are vulnerable due to their social or economic situation."[167]

[58] High-risk AI systems must meet the following requirements: (1) possess a risk management system, (2) continually monitor data, (3) provide technical documentation, (4) maintain records, (5) commit to transparency,

[164] Finlay & Walsh, *supra* note 108.

[165] *Id.*

[166] *Id.*

[167] *Council Publishes Proposed Amendments to Draft AI Regulation*, MCCANN FITZGERALD (Dec. 23, 2021), https://www.mccannfitzgerald.com/knowledge/technology-and-innovation/council-publishes-proposed-amendments-to-draft-ai-regulation [https://perma.cc/ECQ8-VCYE].

and (6) allow for human oversight[168] as well as provide a mechanism for evaluation[169] that can be controlled by humans during processing.[170]

[59] Member states will have to appoint competent national authorities to enforce the regulations at the national level and will be asked to develop rules and policies as well as penalties and a fine structure.[171] In addition, they will be encouraged to provide supervision and testing mechanisms to promote the adoption of AI systems that comply with the regulatory guidelines.[172]

B. The UK's Evolving Approach

[60] The United Kingdom's (UK) government automated its benefit system in 2013 with the twin goals of improving access and reducing

---

[168] *See* Hilary Bonaccorsi et al., *European Commission's Proposed Regulation on Artificial Intelligence: Requirements for High-Risk AI Systems*, JD SUPRA (June 18, 2021), https://www.jdsupra.com/legalnews/european-commission-s-proposed-3823933 [https://perma.cc/4HHA-UMWF].

[169] *See* Chris Kemp, *Five things you should know about the EU's draft AI Regulation*, Kemp IT Law (May 20, 2021), https://www.kempitlaw.com/five-things-you-should-know-about-the-eus-draft-ai-regulation [https://perma.cc/7A5K-3DWY].

[170] *See id.*

[171] Cailean Osborne, *The European Commission's Artificial Intelligence Act highlights the need for an effective AI assurance ecosystem*, GOV.UK: CENTRE FOR DATA ETHICS & INNOVATION BLOG (May 11, 2021), https://cdei.blog.gov.uk/2021/05/11/the-european-commissions-artificial-intelligence-act-highlights-the-need-for-an-effective-ai-assurance-ecosystem/ [https://perma.cc/G6FC-UNAG].

[172] *See id.* ("Member States will be encouraged to launch *AI regulatory sandboxes* to promote the safe testing and adoption of AI systems under the direct guidance and supervision of national competent authorities, with preferential treatment for SMEs and startups to support innovators with fewer resources. Competent authorities should also provide tailored guidance to support SMEs and start-ups to ensure the regulation does not stifle innovation.").

administrative costs.[173] However, the system was severely flawed,[174] and many of the most vulnerable were the most harmed.[175] Given the overall strength of the UK economy at the time,[176] the increase in poverty was surprising.[177] Reform efforts in 2019 aimed to remedy some of the system's

[173] *See UK: Automated Benefits System Failing People in Need*, HUM. RTS. WATCH (Sept. 29, 2020, 12:01 AM), https://www.hrw.org/news/2020/09/29/uk-automated-benefits-system-failing-people-need# [https://perma.cc/66M6-Q5S4] (explaining how Universal Credit merges six social security benefits into a single monthly lump sum and recipients are tasked with applying for and overseeing the benefits online).

[174] *See id.* (explaining that experts have found that the algorithm underlying the program is flawed and the data the government uses to measure changes in people's earnings "only reflects the wages people receive within a calendar month and ignores how frequently people are paid").

[175] *See id.* (explaining that due to the flawed algorithm of the benefits system, many UK residents "go hungry, fall into debt, and experience psychological distress"); *see 1.3 million people in poor mental health who need help with Universal Credit are being "Set Up to Fail,"* DISABILITY RTS. UK (May 26, 2021), https://www.disabilityrightsuk.org/news/2021/may/13-million-people-poor-mental-health-who-need-help-universal-credit-are-being-%E2%80%9Cset [https://perma.cc/QR5B-U9QV](explaining research by the Money and Mental Health Institute (MMHI) indicating that approximately 1.3 million disabled individuals with "high levels of mental distress" may experience difficulty in obtaining the support necessary to obtain their Universal Credit payments).

[176] *See* John McKenna, *Why the UN is investigating poverty in the United Kingdom*, WORLD ECON. F. (Nov. 16, 2018), https://www.weforum.org/agenda/2018/11/united-nations-investigating-poverty-united-kingdom [https://perma.cc/84EA-WA8P] (highlighting that the UK has the fifth largest economy in the world.).

[177] *See id.* (explaining that since 2010, an additional one million children now live in poverty, despite the UK economy having expanded by more than $220 billion within that time); Ryan Shorthouse et al., *Helping Hand? Improving Universal Credit*, BRIGHT BLUE 137-38 (Mar. 2019), https://brightblue.org.uk/wp-content/uploads/2019/03/Helping-hand.pdf [https://perma.cc/5QY2-EJ4G] (highlighting overall, the elderly, individuals with a physical or mental disability, self-employed people, and long-term unemployed people are most likely to experience hardship with UC's design features).

problems.[178] New measures in response to the coronavirus pandemic made positive impacts.[179]

[61] A 2020 UK Report made recommendations to three United Kingdom audiences: the government, regulators, and public bodies.[180] It begins with an evaluation of whether the Nolan Principles of public service could readily be applied to the use of artificial intelligence; the report determined that no reformulation of those principles was required.[181] The Nolan Principles are (1) selflessness, (2) integrity, (3) objectivity, (4) accountability, (5) openness, (6) honesty, and (7) leadership.[182] The American Society of Public Administration's Code of Ethics recognizes similar principles of public service, and thus the 2020 UK Report provides useful guidance in this jurisdiction.[183] Those principles include the following: (1) the importance of "advanc[ing] the public interest, (2)

---

[178] Shorthouse et al., *supra* note 177, at 138–39 (explaining how, in order to minimize the challenges facing claimants, researchers have proposed a set of supplemental policies, which aim to increase personalization and improve progressivity and recommended solutions on how to better the experience of claimants during the UC process).

[179] *See* Serina Sandhu, *Universal Credit and benefits changes in 2020: when new DWP guidelines come into place–and who they affect*, INEWS (Jan. 7, 2020, 5:05 PM), https://inews.co.uk/news/uk/universal-credit-changes-2020-benefits-dwp-guidelines-new-when-date-explained-381252 [https://perma.cc/CQF2-8T2R].

[180] *Artificial Intelligence and Public Standards: A Review by the Committee on Standards in Public Life*, THE COMM. ON STANDARDS IN PUB. LIFE 1, 4 (Feb. 11, 2020), https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/868284/Web_Version_AI_and_Public_Standards.PDF [https://perma.cc/B6MQ-BDHE] [hereinafter *Public Standards*].

[181] *Id.* at 6.

[182] *The Seven Principles of Public Life*, GOV.UK (May 31, 1995), https://www.gov.uk/government/publications/the-7-principles-of-public-life/the-7-principles-of-public-life--2 [https://perma.cc/5ATT-RGNE].

[183] *Code of Ethics*, AM. SOC'Y FOR PUB. ADMIN., aspanet.org/ASPA/Code-of-Ethics/Code-of-Ethics.aspx [https://perma.cc/DLT7-KE4R].

uphold[ing] the Constitution and the [l]aw, (3) promot[ing] democratic participation, (4) strengthen[ing] social equity, (5) fully inform[ing] and [advising], (6) demonstrat[ing] personal integrity, (7) and [p]romot[ing] [e]thical [o]rganizations."[184]

[62] The 2020 UK report noted that the main challenges AI posed were with the principles of openness, based on inadequate governmental information undermining transparency, and objectivity, given the risks of increasing discrimination based on data bias.[185]

[63] On the issue of ethical principles and guidance, the 2020 UK Report explains that there are "three different sets" of principles and frameworks operating and it may be "unclear how [they] work together."[186] Recommendation 10 states that those who provide public services "must consciously tackle issues of bias and discrimination by ensuring they have taken into account a diverse range of behaviours, backgrounds and points of view. They must take into account the full range of diversity of the population and provide a fair and effective service."[187] The report further recommends that public bodies:

> must maximise diversity at all stages of the AI process to help tackle issues of bias and discrimination within AI systems. There needs to be diversity in the workforce and in training and education, so that biases, whether conscious or unconscious, are less likely to be programmed into AI systems. This includes those building and developing AI systems, and those who have responsibility for AI at various stages of deployment. An increased access to a wider range of skills and perspectives at each stage of

[184] *Id.*

[185] *See Public Standards*, *supra* note 180, at 6.

[186] *Id.* at 33.

[187] *Id.* at 9.

> the process will help public bodies to better consider the impact of AI systems on public standards, and to mitigate the risks identified. Datasets used to train machine learning algorithms will also need to be diverse, so that they work accurately and objectively on different individuals and populations.[188]

[64] Chapter 2 of the 2020 UK Report addresses AI and the Nolan Principles, also known as the Seven Principles of Public Life, which are widely accepted in the United Kingdom.[189] On the openness or transparency issue, the 2020 UK Report noted that many indicated it was difficult to even find out where, when, and how the government was using AI, and that most information came from journalists and Freedom of Information Requests (FOIs).[190] For instance, the Guardian reported that almost half of the councils in Great Britain use computer algorithms "despite concerns about their reliability."[191] Thus, the preference was to keep the humans in the loop.[192]

[65] Expanding upon the need for clear guidance in public sector governance, the 2020 UK Report recommends "establishing a new set of

---

[188] *Id.* at 59.

[189] *See id.* at 16-17 (discussing why openness matters and tempers the concept by noting that openness does not require every detail to become public, but that openness is important for sharing the purpose of the technology to the public).

[190] *See Public Standards*, *supra* note 180, at 18.

[191] Sarah Marsh & Niamh McIntyre, *Nearly half of councils in Great Britain use algorithms to help make claims decisions*, THE GUARDIAN (Oct. 28, 2020, 11:00 AM), https://www.theguardian.com/society/2020/oct/28/nearly-half-of-councils-in-great-britain-use-algorithms-to-help-make-claims-decisions [https://perma.cc/E328-UVFP] (noting government use of programs with accuracy rates ranging from 26-40%).

[192] *Public Standards*, *supra* note 180, at 20, 22 ("Many contributors took the view that AI should not retain any role in making a final decision, particularly where the adverse effects on an individual could be significant. They suggested instead that AI should be thought of as a decision-support tool, rather than a decision-making system.").

values and principles, known as the FAST Track Principles and the SUM Values."[193] SUM stands for "'support, underwrite, and motivate a responsible innovation ecosystem' by outlining the values that underpin the ethical permissibility of an AI project. Those values are respect, connect, care, [and] protect."[194] The FAST principles "guide the design and use of AI systems. They are fairness, accountability, sustainability and transparency."[195] Applying the Equality Act of 2010, the 2020 UK Report notes that "there is no reason to view discrimination resulting from biased data differently from discrimination resulting from human bias. Both undermine the Nolan Principle of objectivity."[196]

[66] Governments should also "consider how an AI impact assessment could be integrated"[197] into existing processes to evaluate the potential effects of AI on public standards. Such an assessment should be mandatory and should be published.[198] These assessments will help to redesign or retrain algorithms before they are deployed to minimize unintended or harmful behavior.

[67] For instance, the public sector must "justify why they are using an algorithm; consider whether the potential impact on individuals is necessary and proportionate; and demonstrate how the tool will improve the current system."[199] In addition, "[p]ublic bodies will need to be able to explain and

---

[193] *Id.* at 31.

[194] *Id.*

[195] *Id.* at 31–32 (noting that fairness requires a "minimum threshold of discriminatory non-harm").

[196] *Id.* at 45.

[197] *Public Standards*, *supra* note 180, at 54.

[198] *See id.* at 54; *see also* Andrew Selbst, *An Institutional View of Algorithmic Impact Assessments*, 35 HARV. J.L. & TECH. 117 (2021) (algorithmic impact assessments).

[199] *Public Standards*, *supra* note 180, at 57.

justify decisions made by AI technology. This means that they need to be auditable and transparent enough to satisfy a proper process of appeal and redress."[200] This "proper process of appeal and redress" is analogous to the U.S. Constitution Due Process Clause, and the principles that Judge Friendly announced.[201]

[68] Similar challenges apply in the private sector, but such challenges are beyond the scope of this paper.[202]

V. AREAS FOR FURTHER RESEARCH AND REVIEW

[69] Compiling research from the United States, UK, and EU, we see there are numerous considerations when evaluating potential challenges to government uses of AI. This section lists those challenges impacting fairness and procedural due process. Of course, there are a plethora of issues and questions to address and answer in the expansive world of AI, many of which Lisa Lifshitz and Cameron McMaster have compiled for the journal SciTechLawyer.[203]

[70] Human Agency and Oversight: How does the AI system safeguard against human overconfidence or misuse? Does it unduly restrict human decision-making? Where and when are "humans in the loop?"

---

[200] *Id.* at 65 (explaining that audits are necessary to discover how AI systems work and make decisions because public bodies need to be able to track the process by which a system was designed, procured and deployed, and should be able to trace the way an automated decision was made).

[201] *Id.*

[202] *See* Adam Vaughan, *Companies could be fined if they fail to explain decisions made by AI*, NEW SCIENTIST (Dec. 2, 2019), https://www.newscientist.com/article/2225186-companies-could-be-fined-if-they-fail-to-explain-decisions-made-by-ai [https://perma.cc/696Q-KW46] (declaring that businesses could face multimillion-pound fines if they cannot explain decisions made by AI).

[203] Lisa R. Lifshitz & Cameron McMaster, *Legal and Ethics Checklist for AI Systems*, 17 THESCITECHLAWYER 28, 29–34 (2020).

[71] Accuracy and Reliability: How is accuracy calibrated and measured? How are incorrect outputs identified and addressed? What steps are taken to increase accuracy? How is the reproductivity of outputs evaluated?

[72] Traceability and Explainability: Can we retrace the system design? Can AI system choices be explained in a way that is comprehensible to the government actor and to the impacted member of the public? How is the reasoning process communicated to them?

[73] Bias and Fairness: What is the training dataset? How diverse was the design team? What steps were taken to ensure diversity and representation in the data set? Were bias-preserving measures incorporated into the algorithm? If so, what measures safeguard individuals from unfairly-biased decisions? Is auditing available? How regularly are audits conducted? Can humans identify potentially biased outcomes, and what is the process for addressing them?

VI. CONCLUSION

[74] Answers to the questions posed above will lead to additional follow-up questions tailored to the specific uses of AI. For attorneys considering challenges to government uses of AI, this checklist is a starting point. It is important to not fall into the trap of thinking you understand AI when you do not. The fact that it appears to work does not necessarily imply it is performing the task that has been identified. It is beneficial to recognize and remember that the "status quo is not neutral," and so much of the data AI is trained upon reflects and reinforces the status quo.

[75] New York City has taken a bold step to require algorithms be tested through a bias audit prior to deployment. The parameters of the bias audit are not yet clear, and critics suggest that the passing of the audit will be somewhat of a city endorsement of the AI technologies and whatever biases they may contain, thus leading to an increase in the uses of AI technologies in hiring and screening interviews. The District of Columbia is considering

similar regulation and the EEOC is examining potential Title VII violations when AI technologies are used as well. This issue is evolving quickly.