AI, BIAS, AND NATIONAL SECURITY PROFILING

Laurie N. Hobart†

ABSTRACT

Artificial Intelligence (AI), an increasingly utilized tool for searching, sorting, and analyzing data, has the potential to exacerbate an existing governmental tendency to profile in national security investigations based on ethnicity and 'race,' national origin, and religion. AI is or may be used in national security criminal investigations; intelligence, counterintelligence, or counterterrorism activities; watchlisting practices; border security and immigration investigations; and surveillance programs.

This Article outlines the ways that AI may reproduce human bias in national security investigations at scale. It argues that current case law is insufficient to protect civil rights and civil liberties against discriminatory uses of AI. It discusses potential barriers to constitutional challenges under Fourth Amendment, Equal Protection, First Amendment, and Due Process precedents; limitations on *Bivens* claims; and issues arising from classification, state secrets doctrine, and general judicial deference in national security contexts. While explaining how current case law risks civil rights and civil liberties in the face of AI profiling, the Article also offers litigation strategies for civil rights advocates to proceed under the status quo. First, it argues that the 1996 Supreme Court case *Whren v. United States*, which typically limits profiling challenges to equal protection rather than Fourth Amendment claims, should not apply in the case of AI profiling. Second, it argues that under equal protection law, AI bias should be treated as actionable disparate treatment rather than non-actionable disparate impact.

Executive policies, too, are insufficient to address the potential harms posed by discriminatory AI. The Article addresses helpful and problematic aspects of three executive policies: the 2023 Department of Justice's guidance for law enforcement on the use of suspect categories; the 2020 Artificial Intelligence Ethics Framework for the Intelligence Community, and the 2023 Executive Order on AI and its implementing memoranda. The Article concludes with recommendations for Congress, courts, and government attorneys to guard against national security profiling by humans and machines.

DOI: https://doi.org/10.15779/Z38VX06474

† Associate Teaching Professor, Syracuse University College of Law, Institute for Security Policy and Law (SPL); former Assistant General Counsel within the Intelligence Community. I am grateful to the Hon. James E. Baker and the Georgetown Center for Security and Emerging Technology for the opportunity and funding to study AI; to the AALS Section on National Security for the opportunity to present this work in draft and to Amy Gaudion, Dakota Rudesill, and Alex Sinha and attendees for helpful feedback; to the Syracuse University College of Law Faculty Colloquium organizers and attendees for their helpful comments, especially Kristen Barnes, David Driesen, and Lauryn Gouldin; to SPL research assistants Shannon Cox, Henry DuBeau, Hannah Gabbard, Harrison Gregoire, Thomas Finnigan III, Kaitlyn Keane, Michael Stoianoff, and David Trombly for their excellent research; and to Alyson Chie, Alex Choi, Ryan Hayden, Edlene Miguel, Bani Sapra, Jacob Shofet, Daniel Warner, and other editors of the Berkeley Technology Law Journal for their insightful suggestions and professionalism.

TABLE OF CONTENTS

I. INTRODUCTION

We are a nation that profiles. Not all the time, everywhere, but somewhere, every day. Official profiling actions by our states and country include: the genocide of Native Americans, the taking of their lands and children; the genocide and slavery of Africans, their children and descendants; Jim Crow; the FBI/CIA COINTELPRO operation targeting civil rights leaders; broken window and over-policing of Black Americans; the overincarceration of Black Americans since the abolition of slavery; the exclusion of Chinese immigrants; the detention and family separation of Japanese Americans; McCarthyism, including its targeting of Jewish Americans; the tireless pursuit of Mexican and

other Latin American immigrants, culminating in the border wall and the separation of families and, again, the taking of children; the over-policing of Latin Americans; the rejection and refoulement of asylum applicants; the rounding up and alleged physical and emotional abuse of Muslim and Arab immigrants after 9/11; the photo, video, and mosque crawling surveillance of Muslim and Arab Americans, and the recruitment of community informants; the travel ban against people from Muslim states; and many more examples. Such a shorthand list could never do justice to the many injustices, to the litany of lives changed. It is a litany of loss, to those individuals profiled and persecuted, and to their local and national communities. Our American record is no exception to the patterns of power, fear, and abuse of the "other" stitched across human history. Have we changed, lessened the pattern over time? Perhaps, perhaps not; but with autocracy on the march around the world and autocratic measures and bigotry advocated for openly by politicians at home, we should not dismiss the possibility of our government furthering the worst practices in our own history.

Now enter AI, from stage right, stage left, and even the orchestra pit. Artificial Intelligence (AI) has overwhelmed the modern scene with an omnipresence that will only deepen. We are increasingly aware of the surveillance effects of the Internet of Things, of the constant tracking we submit to by any number of private companies, government agencies, and rogue internet actors. AI tools are being employed in all fields: medicine and health care, online shopping, social media, and environmental protection, to name a few. Generative AI, such as ChatGPT, is poised to change the practice of many disciplines, including law. It may one day help address problems, such as climate change, or create horrors, such as "dangerous biochemicals."[1] AI has the potential to bring great benefits to humanity but also great risks.[2] Among those risks, "algorithmic bias" has been a source of much debate and research. AI developers seek to improve models to mitigate bias, but experts agree that some bias is inherent in AI, just as it is inherent in humans.

Much has been written about AI in the criminal justice system, such as predictive policing algorithms and risk assessments used by courts for bail, parole, and even sentencing decisions.[3] The algorithms are often, if not always,

1. *How Generative Models Could Go Wrong*, ECONOMIST (Apr. 19, 2023), https://www.economist.com/science-and-technology/2023/04/19/how-generative-models-could-go-wrong.

2. *See id.*; Matt O'Brien & Josh Boak, *Biden, Harris Meet with CEOs About AI Risks*, AP NEWS (May 4, 2023), https://apnews.com/article/ai-artificial-intelligence-white-house-harris-578d623e473b0eeb3fa3e4728d7e9868.

3. For an overview of that literature, see *A Letter to the Members of the Criminal Justice Reform Committee of Conference of the Massachusetts Legislature Regarding the Adoption of Actuarial Risk*

trained on historical data that reflect systemic racism in the criminal justice system, and they produce biased, discriminatory results.[4] Scholarly and media attention there is critical. This Article, however, focuses on the potential for bias and for AI profiling by elements of the national security apparatus, where government AI will operate under the further cloak of secrecy and the shield of even more permissive case law. Some of the arguments advanced here, however, apply equally well to routine criminal justice contexts.

AI is an arguably necessary intelligence tool for searching, sorting, and analyzing data and reporting. But it has the potential to exacerbate an existing government tendency to profile in national security investigations based on ethnicity and 'race,' national origin, and religion, and to reproduce that bias at scale. AI is or likely will be used, for example, in national security criminal investigations; foreign intelligence or counterintelligence operations or investigations; watchlisting practices; border policies and customs investigations, and general monitoring or surveillance programs.[5] The government has a pressing need to use AI for at least some national security and intelligence purposes—that is well argued and documented.[6] But the legal

---

Assessment Tools in the Criminal Justice System, MEDIUM (Feb. 9, 2018), https://medium.com/berkman-klein-center/a-letter-to-the-members-of-the-criminal-justice-reform-committee-of-conference-of-the-massachusetts-2911d65969df; *see also* Martha Minow, Jonathan Zittrain & John Bowers, *Technical Flaws of Pretrial Risk Assessments Raise Grave Concerns*, BERKMAN KLEIN CTR. FOR INTERNET & SOC'Y AT HARVARD UNIV. (July 17, 2019), https://cyber.harvard.edu/story/2019-07/technical-flaws-pretrial-risk-assessments-raise-grave-concerns.

4. *See, e.g.*, Minow, et al., *supra* note 3, at 3–4; *The Use of Pretrial "Risk Assessment" Instruments: A Shared Statement of Civil Rights Concerns*, https://civilrightsdocs.info/pdf/criminal-justice/Pretrial-Risk-Assessment-Full.pdf; *Liberty at Risk, Pre-trial Risk Assessment Tools in the U.S.*, EPIC (Sept. 2020), https://epic.org/issues/ai/ai-in-the-criminal-justice-system/; Julia Angwin, Jeff Larson, Surya Mattu & Lauren Kirchner, *Machine Bias*, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

5. For purposes of this Article, I will use the ACLU's definition of profiling as the discriminatory practice by law enforcement (and I will add "by intelligence officials") of targeting individuals for suspicion of crime (and, I will add, "for intelligence interest") based on the individual's race, ethnicity, religion, or national origin. *See What is Racial Profiling?*, ACLU, https://www.aclu.org/issues/racial-justice/race-and-criminal-justice/racial-profiling. This Article does not address the government's national-origin-related categorization of U.S. persons and non-U.S. persons as required by the Foreign Intelligence Surveillance Act (FISA) of 1976 or the FISA Amendments Act of 2008.

6. *See generally* NAT'L SEC. COMM'N ON ARTIFICIAL INTEL., FINAL REPORT 107–19 (2021), https://reports.nscai.gov/final-report/ (discussing why "[i]ntelligence will benefit from rapid adoption of artificial intelligence (AI)-enabled technologies more than any other national security mission."); Corin Stone, *The Integration of Artificial Intelligence in the Intelligence Community*, JOINT PIJIP/TLS RSCH. PAPER SER. 73 (2021), https://digitalcommons.wcl.american.edu/research/73 ("fus[ing] existing and new recommendations

guardrails are shaky, and at some points along the highway, missing altogether. The pace of both national security practice and AI development is very fast, so guardrails are especially needed. AI is inherently risky for civil rights and civil liberties, perhaps insolvably so, and current case law may fail to protect against AI harms.

This Article outlines the ways that AI may exacerbate and reproduce at scale existing bias in national security investigations and surveillance; argues that existing case law is insufficient to protect constitutional rights of equal protection, religious freedom, and due process, and that existing executive policies are likewise inadequate; and suggests litigation strategies for plaintiffs and approaches for courts, Congress, and executive agencies.

Part II will briefly explain from a technical perspective how bias is produced in and reproduced by AI, how bias might alter investigatory and intelligence outcomes, and how AI might expand the net of people under surveillance.

Part III details how existing national security case law is inadequate and even problematic for the protection of civil rights and liberties against harmful uses of AI. It discusses potential barriers to constitutional challenges under Fourth Amendment, Equal Protection, First Amendment, and Due Process precedents; limitations on *Bivens* claims; and issues arising from classification, state secrets doctrine, and general judicial deference in national security contexts. While explaining how current case law risks civil rights and civil liberties in the face of AI profiling, I also seek to provide litigation strategies for civil rights and civil liberties advocates to proceed under the status quo. Courts likewise might adopt such frameworks when applying precedent to biased AI. I argue that the 1996 Supreme Court case *Whren v. United States*,[7] which typically precludes litigants from challenging law enforcement profiling under the Fourth Amendment where there is at least a pretextual non-discriminatory basis for the search or seizure, should not apply to AI-enabled profiling. I also argue that biased AI outcomes should be treated as disparate treatment, rather than simply disparate impact, and therefore actionable under current Equal Protection law.

Part IV discusses three recent executive policies: the Department of Justice "Guidance for Federal Law Enforcement Agencies Regarding the Use of Race, Ethnicity, Gender, National Origin, Religion, Sexual Orientation, Gender

to prioritize actions that will drive faster, more effective transition of cutting-edge AI into the [Intelligence Community]," while recognizing that questions about ethics, including bias, privacy, transparency and security remain paramount).

7. Whren v. United States, 517 U.S. 806 (1996).

Identity, and Disability,"[8] the Intelligence Community's Artificial Intelligence Principles and Ethics Framework,[9] and the October 2023 executive order on AI[10] and its implementing memoranda. (The 2023 executive order was just revoked by President Trump in January 2025, and policy memoranda directed by it are now under review by his administration, as discussed in Section IV.C.) Collectively, these executive policies have both problematic and helpful aspects for AI governance. While the government has demonstrated technical sophistication in its understanding of AI and, at least in some instances, a dedication to testing AI for bias, any progressive regulations and policies are reversible by future administrations, a process that seems to be underway as of this writing.

Part V suggests solutions that civil rights and civil liberties advocates might pursue in legislation. It also provides recommendations for courts and government attorneys seeking to minimize algorithmic discrimination and national security profiling.

II. HOW BIAS IS PRODUCED IN AI AND REPRODUCED AT SCALE

This Part explains how human bias is reproduced by AI, how such bias might alter investigatory and intelligence outcomes, especially by exaggerating pre-existing human biases within the national security field, and how AI might expand the net of people scoped into surveillance.

A. BIAS IN AI GENERALLY

1. Algorithmic Bias Defined

Algorithmic bias is a broad term that refers to any difference between the desired accuracy of an AI model and the actual output. This difference might be seen, hypothetically, when an AI system designed to recognize an enemy

8. U.S. DEP'T OF JUST., GUIDANCE FOR FEDERAL LAW ENFORCEMENT AGENCIES REGARDING THE USE OF RACE, ETHNICITY, GENDER, NATIONAL ORIGIN, RELIGION, SEXUAL ORIENTATION, GENDER IDENTITY, AND DISABILITY (May 2023), https://www.dhs.gov/sites/default/files/202306/Guidance%20for%20Federal%20LEAs%20on%20the%20Use%20of%20Protected%20Characteristics_FINAL%205.25.23_508.pdf.

9. OFF. OF THE DIR. OF NAT'L INTEL., ARTIFICIAL INTELLIGENCE ETHICS FRAMEWORK FOR THE INTELLIGENCE COMMUNITY, VERSION 1.0, https://www.intelligence.gov/artificial-intelligence-ethics-framework-for-the-intelligence-community (last visited Feb. 4, 2023).

10. Exec. Order No. 14,110, 88 Fed. Reg. 75,191 (Oct. 30, 2023). This Order was revoked by President Trump on January 23, 2025.

tank at night, trained on images of tanks at night, mistakes, in the dark, the box-like shape of an air conditioner unit on a rooftop for a tank.[11]

Algorithmic bias includes under its umbrella discriminatory bias, where the AI model discriminates against or disparately impacts a group of people on the basis of their 'race,' ethnicity, gender, religion, national origin, disability, sexuality, or other axis of differentiation. Discriminatory bias might determine, for example, who is being delayed or detained at the airport or border. The Biden Administration's 2023 executive order on AI and the Draft AI Bill of Rights[12] helpfully use the term "algorithmic discrimination" when referring to this type of bias in AI.

There is often an ethical and legal obligation to mitigate bias of any kind in AI models: in the tank-recognition hypothetical, to comply with the law of armed conflict and human rights law; in the case of algorithmic discrimination, to comply with constitutional principles and values. I write "mitigate" because experts agree that there is no way to eliminate bias from any AI model. There is, however, always the option to forgo using AI for a particular task or problem. As Dr. Kush Varshney writes, "In some cases, the problem should not even be solved to begin with [using AI], because doing so may cause or exacerbate societal harms and breach the lines of ethical behavior."[13]

2. How Bias is Produced Throughout the Lifecyle of an AI Model

Bias can enter an AI model in myriad ways. Here, I will discuss some of the most common sources of bias, but this is not an exhaustive list or treatment.[14] This discussion will follow the stages of the machine learning[15]

11. SYRACUSE UNIV. INST. FOR SEC. POL'Y AND LAW & GEORGETOWN CTR. FOR SEC. AND EMERGING TECH., NAT'L SEC. L. & THE COMING AI REVOLUTION 16 (2021), https://cset.georgetown.edu/wp-content/uploads/Symposium-Report-National-Security-Law-and-the-Coming-AI-Revolution.pdf.

12. Exec. Order No. 14,110, 88 Fed. Reg. 75,191, 75,211 (Oct. 30, 2023); WHITE HOUSE OFF. OF SCIENCE & TECH. POL'Y, BLUEPRINT FOR AN AI BILL OF RIGHTS 23–29 (2022).

13. KUSH R. VARSHNEY, TRUSTWORTHY MACHINE LEARNING 16 (2022), *available at* http://www.trustworthymachinelearning.com/trustworthymachinelearning.pdf. Dr. Varshney, an IBM Fellow, directs IBM's Human-Centered Artificial Intelligence and Trustworthy Machine Intelligence teams and was a founding co-director of the IBM Science for Social Good initiative.

14. *See, e.g.*, UNITED NATIONS INST. FOR DISARMAMENT RSCH., UNIDIR RESOURCES NO. 9, ALGORITHMIC BIAS AND THE WEAPONIZATION OF INCREASINGLY AUTONOMOUS TECHNOLOGIES (2018), http://www.unidir.ch/files/publications/pdfs/algorithmic-bias-and-the-weaponization-of-increasingly-autonomous-technologies-en-720.pdf.

15. Machine learning is one of the main fields of artificial intelligence; most models today, from predictive shopping algorithms to ChatGPT, use machine learning.

(ML) lifecycle—that is, the lifecycle of an AI model.[16] Lawyers need to know that human judgement, values, and biases are inserted into AI models at every step of that process.[17] Moreover, every AI model is unique, and many models evolve and learn (for better or for worse) over the course of their "lives," from their conception and design through their training and deployment in the field through their dying days (end of use). Bias is a moving target.

Bias can be produced by the original framing of the question or task posed to the AI to answer or perform.[18] Designing a facial recognition system to unlock a phone has a different ethical implication from creating one to identify and track a particular population, such as the Uighur people in China.[19] (And in between those two examples there is still much room for harm to particular groups.) Human values, biases, preferences, and judgments inform any question presented to an algorithm, just as they inform any question presented to a court in a legal brief. The perhaps ultimate ethical and legal question first occurs at this origin point in the ML lifecycle: should the AI be developed and used for a particular purpose? Or does that very use risk too much vis-à-vis civil rights and civil liberties at home or human rights around the world? Experts suggest consulting stakeholders, including groups most likely to be affected by the AI, especially marginalized groups.[20] The question of whether

16. Varshney, *supra* note 13, at 14–22 (citing the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology); *see also* Nick Hotz, *What is the Data Science Process?*, DATA SCI. PROCESS ALL., https://www.datascience-pm.com/data-science-process/ (last updated Nov. 18, 2024) (identifying the CRISP-DM and other life cycle models for data science).

17. DAVID LESLIE, THE ALAN TURING INST., UNDERSTANDING ARTIFICIAL INTELLIGENCE ETHICS AND SAFETY 16 (2019), https://www.turing.ac.uk/sites/default/files/2019-06/understanding_artificial_intelligence_ethics_and_safety.pdf.

18. *Id.*

19. *See, e.g.*, Paul Mozur, *One Month, 500,000 Face Scans: How China Is Using A.I. to Profile a Minority*, N.Y. TIMES (Apr. 14, 2019), https://www.nytimes.com/2019/04/14/technology/china-surveillance-artificial-intelligence-racial-profiling.html; Manoush Zomorodi, Katie Monteleone & Sanaz Meshkinpour, *How Facial Recognition Allowed the Chinese Government to Target Minority Groups*, NPR (Dec. 9, 2022), https://www.npr.org/2022/12/09/1141627539/how-facial-recognition-allowed-the-chinese-government-to-target-minority-groups.

20. Varshney, *supra* note 13, at 15, 17 (citing Meg Young, Lassana Magassa & Batya Friedman, *Toward Inclusive Tech Policy Design: A Method for Underrepresented Voices to Strengthen Tech Policy Documents*, 21 ETHICS AND INFO. TECH. 89 (2019)); Leslie, *supra* note 17, at 16; *see also* OFF. OF THE DIR. OF NAT'L INTEL., ARTIFICIAL INTELLIGENCE ETHICS FRAMEWORK FOR THE INTELLIGENCE COMMUNITY, VERSION 1.0, https://www.intelligence.gov/artificial-intelligence-ethics-framework-for-the-intelligence-community (last visited Sept. 7, 2023) ("Identifying and addressing risk is best achieved by involving appropriate stakeholders. As such, consumers, technologists, developers, mission personnel, risk management professionals, civil liberties and privacy officers, and legal counsel should utilize this

to use or forgo using the AI must be repeated throughout its lifecycle, as it or real-world conditions change. Likewise, consultation with stakeholders is critical throughout the machine learning lifecycle.[21]

Human values and biases also inform the metrics by which we measure the AI's accuracy or success:[22] Is the algorithm successful if it correctly identifies terrorism suspects? Or is it successful only if it correctly identifies suspects without also incorrectly flagging innocents? Or only if it correctly identifies suspects based on their realized, individual behavior (such as purchasing bomb-making equipment or a particular weapon) without also incorrectly flagging innocents or disproportionately flagging certain groups of people? And so forth. One can build as many (or as few) qualifiers as one chooses into the metrics, always keeping in mind the initial question of whether the potential societal harms are too great to make success by any metric worthwhile.

Bias might also result from the data on which an AI is trained, tested in the lab, or later validated in the field. Biased inputs produce biased outputs (the oft-quoted "garbage in, garbage out"). Two common sources of bias are "temporal bias," where the AI is trained on historical data that may not reflect current societal values[23] (voting data, for example, where certain groups were disenfranchised), and "population bias" or "representation bias" where some groups are over- or under-represented in the data.[24] Representation bias can also include differences in quality or labeling or engineering of data by technologist across groups.[25] Algorithms used for predictive policing, bail, parole, and sentencing in the criminal justice system likely reflect both temporal and population bias. One of the main criticisms of the use of AI by police and the courts is that the datasets upon the AI are trained reflect the *historically disproportionate* policing, arrest, prosecution, and sentencing of people of color for the same conduct as white people.[26] People of color are overrepresented in the training and possibly validation and testing data sets, while white people are underrepresented. Another example of population bias is that early (and some current) facial recognition algorithms performed

framework collaboratively, each leveraging their respective experiences, perspectives, and professional skills").

21. *Id.*

22. *See* Nisheeth Vishnoi, A. Bartlett Giamatti Professor of Comput. Sci., Yale Sch. of Eng'g and Applied Sci., Remarks at the Yale Cyber Leadership Forum (Feb. 18, 2022), https://youtu.be/VT-j_YgODUw?si=Wu7BQXHjrLMERbfI.

23. *See* Varshney, *supra* note 13, at 48.

24. *Id.*

25. *Id.*

26. *Technical Flaws of Pretrial Risk Assessments Raise Grave Concerns*, BERKMAN KLEIN CTR. FOR INTERNET & SOC'Y AT HARVARD UNIV. (July 17, 2019), https://cyber.harvard.edu/story/2019-07/technical-flaws-pretrial-risk-assessments-raise-grave-concerns.

especially poorly at accurately matching faces of women of color because similar faces were underrepresented in the training data.[27]

Proxies in data are another source of bias. Proxies are pieces of data that might correlate with another type of information, such as a legally protected classification. For example, zip codes might serve as proxies for demographic information (racial, ethnic, etc.) or, in some cultures, last names for religion.[28] Housing and employment data used in criminal risk assessments have proved to be strong proxies for 'race' and class.[29] An algorithm may not explicitly be programmed to make such a connection but nonetheless learn to do so.

How data scientists and engineers label, clean, and manipulate data will also increase (or potentially decrease) the amount or type of bias. "Cleaning" data involves, among other things, filling in missing values or discarding them, binning continuous feature values to account for outliers, grouping or recoding features, and dropping features, including features that "should not be used for legal, ethical, or privacy reasons."[30] Feature engineering data involves "mathematically transforming features to derive new features" and requires real "creativity from data scientists."[31] It is, in other words, another entry point for human skill, error, and bias.

Once the data is prepared, engineers select and develop an algorithm. An algorithm, definitionally, is a set of instructions, such as a recipe; in the AI world, it is a set of mathematical instructions. Engineers select the inputs (what features, expressed in numbers, the model will evaluate or process), all the parameters or connections between inputs and later interpretations of the inputs, and the weight assigned to each input.

Many models today are "closed boxes" or "black boxes"—that is, engineers know the inputs fed to the algorithm and the outputs it produces but not how or why it produces the outputs that it did.[32] This lack of

27. *See* Tonya Mosley, *'If You Have a Face, You Have a Place in the Conversation about AI,' Expert Says*, NPR (Nov. 28, 2023), https://www.npr.org/2023/11/28/1215529902/unmasking-ai-facial-recognition-technology-joy-buolamwini (interviewing Dr. Joy Buolamwini).

28. *See* Varshney, *supra* note 13, at 18.

29. Chelsea Barabas, Christopher T. Bavitz, Ryan H. Budish, Karthik Dinaker, Cynthia Dwork, Urs Gasser, Kira Hessekiel, Joichi Ito, Ronald L. Rivest, Madars Virza & Jonathan Zittrain, *Open Letter to the Members of the Massachusetts Legislature Regarding the Adoption of Actuarial Risk Assessment Tools in the Criminal Justice System*, BERKMAN KLEIN CTR. FOR INTERNET & SOC'Y AT HARVARD UNIV. 3 (Nov. 9, 2017) http://nrs.harvard.edu/urn-3:HUL.InstRepos:34372582 (citing Sonja B. Starr, *Evidence-Based Sentencing and the Scientific Rationalization of Discrimination*, 66 STAN. L. REV. 803 (2014)).

30. Varshney, *supra* note 13, at 19.

31. *Id.*

32. *Id.* at 21.

transparency poses significant due process and equal protection issues. There are, however, other methodologies available; there is increasing research on more transparent neural networks that will allow more insight into the decisions, weights, and parameters.[33]

A model is evaluated or tested on a separate, held-out set of "validation" data.[34] The model is fed inputs from the validation data set to see whether and how well the model works. Validation data can also be a source of bias as the AI learns and potentially readjusts its algorithms based on that data.

Finally, the AI is deployed in the field, where it interprets real world data. In the field, the model should be constantly monitored because its accuracy and trustworthiness can "degrade over time" as it incurs real world inputs that drift from training data.[35] For example, in 2016, the chatbot "Tay" generated and spread hate speech within hours of its public release because it adopted the language and values found on Twitter.[36]

An AI model will reflect historical or current biases of its consumers, designers, engineers, its training and validation data, and the real-world data and users it encounters when deployed. Every decision point—and individual life—the AI touches in its lifetime will be affected by those original sins.

3. Unconscious, Culturally Produced Bias

The human bias that is built into the AI might be explicit or implicit, intentional or unconscious.[37] It may or may not have been *consciously* intended by the designers. As Professor Charles Lawrence argued in 1987, because racism is part of the American historical and cultural heritage, we "inevitably share many ideas, attitudes, and beliefs that attach significance to an individual's race and induce negative feelings and opinions about nonwhites."[38] Racism is often unconscious because we do not recognize how "our cultural experience has influenced our beliefs about race," nor "the occasions on which

33. *See id.*; *see also AI Fundamental Research – Explainability*, NAT'L INST. OF STANDARDS AND TECH. (June 16, 2022), https://www.nist.gov/artificial-intelligence/ai-fundamental-research-explainability.

34. Varshney, *supra* note 13, at 21.

35. *Id.*

36. Oscar Schwartz, *In 2016, Microsoft's Racist Chatbot Revealed the Dangers of Online Conversation*, IEEE SPECTRUM (Nov. 25, 2019), https://spectrum.ieee.org/in-2016-microsofts-racist-chatbot-revealed-the-dangers-of-online-conversation.

37. Nonetheless, as a I argue below, the government should be regularly evaluating AI models for bias, and if the government chooses to use a biased AI model, that choice may be considered an intentionally discriminatory act under current case law.

38. Charles Lawrence, *The Id, the Ego, and Equal Protection: Reckoning with Unconscious Racism*, 39 STAN. L. REV. 317, 322 (1987).

those beliefs affect our actions."[39] Professor Lawrence went on to explain how Freudian and cognitive theory each suggest that we unconsciously adopt our culture's beliefs and preferences, even those we consciously judge as immoral.[40] Since that writing, there has been much research on social and cognitive biases,[41] all important to understanding how we may program bias into AI. Of central importance to Professor Lawrence, as he later wrote, however, was "the cultural meaning of racial texts";[42] he sought to explore "how white supremacy is maintained not only through the intentional deployment of coercive power, but also through the creation, interpretation, and assimilation of racial text."[43] When humans program AI, machines will read our cultural texts, with all of the biases and prejudices they contain, and implement and spread those ills.

Intelligence and law enforcement officials are influenced, in many cases unconsciously, not only by the national culture, but also by national-security-community and agency-specific cultures.[44] Historically, the intelligence and law enforcement communities have targeted the civil rights movement, the anti-war movement, Muslims, and more. One of the reasons it is so important to diversify the Intelligence Community (IC) and law enforcement is that a body of civil servants with diverse experiences and cultural influences will produce more creative, perceptive, and reliable intelligence results. But diversity work is far from complete in the IC[45] and more broadly under attack by the Trump Administration.[46] Nor can diversity efforts be the only solution to the problems of bias and limited perspectives in intelligence activities.

Professor Lawrence sought "to advance the understanding of racism as a societal disease and to argue that the Constitution commands our collective responsibility for its cure."[47] As explored in the next section, we should expect that current biases in the national security field and particular agencies may be reflected in the development of or the choice to use particular AI models or

39. *Id.*

40. *Id.*

41. *See* Charles Lawrence, *Unconscious Racism Revisited: Reflections on the Impact and Origins of "The Id, the Ego, and Equal Protection,"* 40 CONN. L. REV. 931 (2008).

42. *Id.* at 938–39.

43. *Id.* at 939.

44. *See* Shirin Sinnar, *Separate and Unequal: The Law of "Domestic" and "International" Terrorism*, 117 MICH. L. REV. 1333, 1388 n.307 (2019).

45. *See, e.g.*, GAO-21-83, *Intelligence Community: Additional Actions Needed to Strengthen Workforce Diversity Planning and Oversight*, U.S. GOV'T ACCOUNTABILITY OFF. (Dec. 17, 2020), https://www.gao.gov/products/gao-21-83.

46. *See, e.g.*, Exec. Order 14151, Ending Radical and Wasteful Government DEI Programs and Preferencing, 90 Fed. Reg. 8,339 (2025).

47. Charles Lawrence, *Unconscious Racism Revisited*, *supra* note 41, at 942.

AI for particular purposes. We have a collective, constitutional duty to counter the cultural racisms and other biases that AI will invariably adopt and disseminate, and to redress AI harms.

B. AI AS A POTENTIAL HUMAN-BIAS MULTIPLIER IN NATIONAL SECURITY CONTEXTS

It is too easy to imagine parallels to AI biases in the criminal justice system playing out in the national security context. Most national security AI programs are likely classified, so to some extent we really must imagine the risks and harms. Here, however, are some possibilities for how AI might exacerbate bias in national security applications.

1. Encoding Historical and Systemic Biases

As AI encodes historical and systemic biases, certain groups may be flagged more often for greater surveillance or negative attention from law enforcement and intelligence officials. The AI might be trained on data from watchlisting, for example, where the U.S. government has historically disproportionately included Muslims and Muslim Americans.[48] Or AI might be trained on data about from events labeled "international" terrorism, and not from events labeled "domestic" terrorism. But as Professor Shirin Sinnar has demonstrated, our legal architecture has created a false dichotomy between international and domestic terrorism.[49] For example, it often categorizes threats of terrorism by Muslim Americans as international, even when there is no evidence of international ties, and threats by white supremacist or neo-Nazi Americans as domestic terrorism, even though they might be influenced by the global supremacy movement.[50] This false dichotomy poses significant harms to Muslim Americans, such as harsher sentencing, disproportionate surveillance of their communities,[51] and "distorted public perceptions of terrorism that fuel anti-immigrant and discriminatory policies."[52] Professor Sinnar also suggests that the category of "international" terrorism "will

48. *See* Letter to Executive Officials from 13 Senators and Members of Congress regarding the Terrorist Screening Dataset (TSDS, or "terrorist watchlist") (Dec. 20, 2023), https://www.warren.senate.gov/imo/media/doc/2023.12.20%20Terrorism%20Watchlist%20Letter.pdf ("Muslim Americans disproportionately face the risk of being wrongfully placed on the watchlist. Advocates have estimated that as many as 98% of people on the list are Muslim, and no evidence that the person has committed or will commit a crime is required, leading some to refer to the watchlist system as a "Muslim registry.").

49. *See generally* Sinnar, *supra* note 44, at 1337.

50. *Id.*

51. *Id.*

52. *Id.* at 1333.

predictably expand to cover U.S. individuals perceived as 'foreign,' even if they are citizens with negligible relationships abroad."[53] Significantly, the false dichotomy may cause us to underestimate the risk of violent harm by "domestic" terrorists.[54] AI that adopts the false dichotomy will be inaccurate and increase all of those harms.

Moreover, any integrated government systems reliant on the outputs produced by the biased AI will also be infected by the same biases. The potential for harm—and biased harm—is evidenced by analogy in watchlisting cases, even without AI involved. In *Ibrahim v. Department of Homeland Security*,[55] one government employee's mistake infected multiple watchlists. In *Elhady v. Kable*,[56] the Eastern District of Virginia, though later reversed on other grounds, determined that the Terrorist Screening Center's Terrorist Screening Database (TSDB), the central national database from which all other, shorter lists are derived, posed due process issues. Those due process issues included the low standard—the executive's reasonable suspicion standard—for inclusion on the lists. Plaintiffs alleged that the Terrorist Screening Center might consider 'race,' ethnicity, religious affiliation and beliefs, and other First Amendment activities in adding individuals to the database.[57] TSDB data is shared with federal, state, local, and tribal partners, and more than sixty foreign governments,[58] compounding the due process harm of any error.[59] Analogously, if biased AI causes an individual or their data to be added to a government system or list, that mistake might proliferate across other connected systems. AI's implications for government watchlists are addressed in further detail in Section III.D, below.

2. Government Aggregation of Data Types and Systems

If various government agencies and private entities merge and mine the many types of data they have collected, that will cause greater privacy

53. *Id.*

54. *See id.* at 1388–92.

55. Ibrahim v. Dep't of Homeland Sec., 62 F. Supp. 3d 909, 916 (N.D. Cal. 2014).

56. Elhady v. Kable, 391 F. Supp. 3d 562, 573 (E.D. Va. 2019) (*reversed by* Elhady v. Kable, 993 F.3d 208 (4th Cir. 2021)). 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/.

57. Elhady v. Kable, 391 F. Supp. 3d at 569.

58. *Id.* at 569–70.

59. *See id.* at 580.

invasion[60] and harms from any errors. Data collections might be combined and analyzed across types (e.g., biometric information with tax information with financial records with criminal records, etc.) and across collecting entities (e.g., law enforcement with intelligence, or federal with state and local). With respect to biometric data, for example, Professor Margaret Hu has examined "the merger of civilian and military, along with domestic and foreign mass biometric data harvesting" and "the potential long-term cybersurveillance consequences of the increased sharing of biometric databases between military, intelligence, and law enforcement organizations, and other public and private entities."[61] She argues that "biometric cybersurveillance and biometric cyberintelligence objectives are increasingly used to justify the mass digital capture and analysis of unique physiological and behavioral traits of entire populations and subpopulations."[62] Data aggregation and harvesting, of course, can be enabled by AI. And biases and biased mistakes present in shared databases and systems will be amplified across them.

One great irony of AI is that while it may help us to organize and analyze the seemingly infinite amount of information in the world—it seems to offer the solution to the office, library, or national government system full of records upon records—it will also create new organizational challenges. How can we trace or root out the bias, mistakes, and even hallucinations[63] that will infiltrate and reproduce in linked AI datasets and systems? It is a Herculean task, perhaps one to which AI itself can be applied, but a challenge of AI-proportions, nonetheless.

Whether and to what extent the government aggregates different types of data is unknown, but absent federal and state[64] legislation there is little to protect against the privacy invasion of aggregated data beyond the application of the Fourth Amendment. Like Congress, courts have been reticent to

60. *See* STEPHEN DYCUS, WILLIAMS C. BANKS, EMILY BERMAN, PETER RAVEN-HANSEN & STEPHEN I. VLADECK, NATIONAL SECURITY LAW 739 (7th ed. 2020) (citing Daniel J. Solove, *A Taxonomy of Privacy*, 154 U. PA. L. REV. 477, 507 (2006)).

61. Margaret Hu, *Biometric Cyberintelligence and the Posse Comitatus Act*, 66 EMORY L.J. 697, 700 (2017).

62. *Id.* at 700–01.

63. Hallucinations are a phenomenon associated with generative AI, specifically large language models (LLMs). A generative AI model is one that creates content, such as text, art, or music. LLMs like ChatGPT generate language responses to user's prompts. The LLM "hallucinates" when it creates inaccurate or misleading outputs that appear authentic. *What are AI Hallucinations?*, IBM, https://www.ibm.com/topics/ai-hallucinations; *When AI Gets It Wrong: Addressing AI Hallucinations and Bias*, MIT MGMT. STS TEACHING & LEARNING TECHNOLOGIES, https://mitsloanedtech.mit.edu/ai/basics/addressing-ai-hallucinations-and-bias/.

64. Some states, such as Illinois and California, have made strides at legislation.

articulate the ground rules. As discussed below, the 2018 case *Carpenter v. United States*, where the Supreme Court held that police needed a warrant to obtain 127 days' worth of cell site data, provides some hint that the Constitution might require the government to seek a warrant before aggregating an individual's data. However, the Court specifically carved national security applications out of its holding.

3. How AI Expands the Net of Surveillance

AI-enabled sensors and systems will also expand the net of surveillance thrown over Americans and people around the globe. AI enables, among other things, facial recognition and other biometric analyses, remote cameras and drones, data aggregation, data mining, and link analysis. In effect, AI has the potential to create a system of what Chief Justice Roberts might recognize as "near perfect surveillance," as he described cell phone location tracking in *Carpenter*.[65] AI-enabled surveillance occurs or may occur not only in public spaces but also in our homes and offices,[66] via our computers, phones, personal electronic assistants, wearable heath monitors, connected appliances, cars, and sound and security systems. As was demonstrated in *Carpenter*, historical data about an individual may be preserved for years before retroactively being searched by the government. The government (and private actors) may also have access to real-time data from live video and sound surveillance, perhaps enhanced by AI-enabled facial, voice, gait, or other biometric-recognition. Such surveillance potential creates tremendous First and Fourth Amendment issues, the subject of much scholarship.[67]

AI increases the scope of government investigations, enabling surveillance of many more people than traditional "gumshoe" police work.[68] Any

65. Carpenter v. United States, 585 U.S. 296, 312 (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.").

66. There is a growing body of literature on AI in the workplace. *See, e.g.*, Karen E. C. Levy, *The Contexts of Control: Information, Power, and Truck-Driving Work*, 31 INFO. SOC'Y 160 (2015), https://www.tandfonline.com/doi/full/10.1080/01972243.2015.998105 (last accessed Oct. 27, 2019).

67. *See, e.g.*, Margot E. Kaminski, *Regulating Real-World Surveillance*, 90 WASH. L. REV. 1113 (2015); Jennifer Lynch, *Face Off: Law Enforcement Use of Face Recognition Technology*, ELEC. FRONTIER FOUND. 1, 8–10 (Feb. 12, 2018), https://www.eff.org/wp/law-enforcement-use-face-recognition#_idTextAnchor004; Claire Garvie, Alvaro M. Bedoya & Jonathan Frankle, *The Perpetual Lineup: Unregulated Police Face Recognition in America*, GEORGETOWN L. CTR. ON PRIV. & TECH. (Oct. 16, 2016), https://www.perpetuallineup.org/.

68. Justice Alito made a related point with respect to how GPS technology increased surveillance capacity against any one person in his concurrence in *United States v. Jones*, 565 U.S. 400, 429 (2012).

discriminatory bias will likewise reach and harm more people. Facial recognition systems offer an example of this expanding ripple effect. Many facial recognition systems, including that of the Federal Bureau of Investigation (FBI), do not necessarily make exact matches. Rather, given a fixed data set, they determine and rank which photos within that set are most likely to match.[69] Facial recognition might flag someone who was five states away from the crime scene; as Electronic Frontier Foundation General Counsel Jennifer Lynch argues, in making that person (or multiple persons) subject to a probable-cause search warrant, we shift the burden from the state proving guilt to the suspect proving his or her innocence: "False positives can alter the traditional presumption of innocence in criminal cases by placing more of a burden on suspects and defendants to show they are not who the system identifies them to be."[70] Moreover, she suggests, biased AI models and databases mean that the harm of false positives will be disproportionately felt by people of color.[71]

If national security AI systems (some of which may overlap with the law enforcement systems Lynch discusses) are biased against people of particular ethnicities, religions, national origins, or other classifications, those groups will disproportionately be ensnared in the net and feel the harm of false positives. For example, a biased AI model used to predict who is most likely to commit a terrorist act might falsely suggest members of certain groups disproportionately more often.

Additionally, as argued below, by expanding the net of people pulled into any investigation, AI may blur the line between using a social identity descriptor for individual suspect identification and using it for group profiling. To take a hypothetical derived from the Department of Justice Guidance,[72] if the government has a tip from a reliable source that an assassin of X descent or nationality is entering the country to kill a diplomat, and the government

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

70. *Id.*

71. *Id.* ("The false-positive risks . . . will likely disproportionately impact African Americans and other people of color. Research—including research jointly conducted by one of FBI's senior photographic technologists—found that face recognition misidentified African Americans and ethnic minorities, young people, and women at higher rates than whites, older people, and men, respectively. Due to years of well-documented racially-biased police practices, all criminal databases—including mugshot databases—include a disproportionate number of African Americans, Latinos, and immigrants. These two facts mean people of color will likely shoulder exponentially more of the burden of face recognition inaccuracies than whites.").

72. U.S. DEP'T OF JUSTICE, *Guidance*, *supra* note 8.

uses AI to search the real-time, AI-enabled video footage of every domestic international airport for people of X descent, then pulls over many of them for additional questioning, that looks like profiling. The greater the number of people swept into the search, the less likely it is that any of those people is the criminal. But all of those people will have their liberty and privacy diminished, without an individually-based predicate for the search or seizure, but rather, based on their ethnicity or national origin. (Of course, at international airports, under border search doctrine there is no Fourth Amendment predicate of reasonable suspicion or probable cause needed for routine searches, but that does not excuse profiling and violating equal protection, nor unreasonable searches.)

4. The Potential Scope of AI National Security Surveillance

Government national security AI uses are largely unknown to the public and likely often maintained as classified.

Professor Emily Berman provides a helpful survey of some reported uses of machine learning by the government in the national security context. AI is used to: help populate the No Fly List; to predict crimes, terrorist acts, and social upheaval; predict whether individuals might be 'mobilized' to violence or terrorism; generate risk-assessments for cargo, vehicles, and individuals at the border; search and analyze federal, state, and local law enforcement databases; data mine the activities and associations of foreigners and Americans; use link analysis to find individuals with connections to terrorism suspects; aid in drone targeting; and analyze reams of aerial surveillance footage.[73]

A Biden Administration policy[74] (currently subject to review by the Trump Administration, as detailed in Section IV.C, below), provides some insight into government uses of AI for national security purposes. Under the policy, agencies must publicly disclose only broad categories of national security AI uses that might have a "high impact" on national security, civil rights, civil liberties, human rights, or democratic norms.[75] Those categories include any AI that "controls or significantly influences the outcomes" of such activities as tracking or identifying individuals in real time based solely on biometric data; classifying an individual as a known or suspected terrorist

73. Emily Berman, *A Government of Laws and Not of Machines*, 98 B.U. L. REV. 1277, 1290–1301 (2018) (internal citations omitted).

74. THE WHITE HOUSE, FRAMEWORK TO ADVANCE AI GOVERNANCE AND RISK MANAGEMENT IN NATIONAL SECURITY ("NS AI Framework") 2–5 (Oct. 24, 2024), https://ai.gov/wp-content/uploads/2024/10/NSM-Framework-to-Advance-AI-Governance-and-Risk-Management-in-National-Security.pdf.

75. *Id.* at 3–4.

or national security threat; determining an individual's immigration or asylum status; and performing criminal risk assessments and predictions inside the United States or with respect to U.S. persons, or in relation to immigration or entry into the United States.[76] Agencies are prohibited from using AI "with the intent or purpose" to profile, target, or track individuals "based solely" on their First Amendment activities, or "[u]nlawfully disadvantage" individuals based on their ethnicity, national origin, race, sex, gender, gender identity, sexual orientation, disability status, or religion.[77] The qualifiers "with the intent or purpose," "based solely," and "unlawfully" significantly limit the scope of those two prohibited categories.

One way to think about the potential scope of classified AI surveillance is to consider that the government could, with only the Constitution standing in its way (i.e., no AI-specific statutory guidance), do anything that China and more authoritarian regimes do with AI. "Only the Constitution standing in the way" might sound like a gross understatement of the very significant protections that document provides. One certainly hopes that government lawyers will carefully consider the constitutional restraints before advising on any proposed AI surveillance programs. However, while government officials might be legally and ethically bound by the Constitution, current case law is not especially helpful in establishing accountability for violations. Government officers—lawyers and the policy officials they advise—typically practice in secrecy. Secrecy makes accountability more difficult; lawyers cannot lean on the threat of liability to encourage policymakers to follow the law. And it might lead some lawyers to advise that there is little litigation risk, which is perhaps factually true but, in my and many practitioners' view, ethically unsound legal advice where constitutional rights stand to be violated. Even if the courts cannot (or choose not to) enforce certain constitutional remedies, government officials are still bound by the Constitution's mandates, most especially the Bill of Rights.

III. CASE LAW IS TOO PERMISSIVE OF POTENTIAL AI PROFILING

As scholars and practitioners know, it is especially challenging for plaintiffs to win cases for constitutional violations by government actors and programs in the national security context. In this section, I will first address how biased

76. *Id.* at 3–4. Department Heads are required to "maintain unclassified public lists of prohibited and high-impact AI categories they have added to these lists, as well as additional categories they have created for additional oversight and safeguards," but their lists "may have classified annexes." *Id.* at 4–5.

77. *Id.* at 3.

AI programs might be adjudicated on the merits, under Fourth Amendment, Equal Protection, First Amendment, and Due Process precedents. While detailing the ways in which current case law may not be protective of civil rights and civil liberties against AI profiling, I also seek to provide litigation strategies for civil rights and civil liberties advocates to proceed under the status quo. Central to those strategies is the idea that AI, though biased, has objectively measurable results. Any time the government adopts a biased AI model, it does so knowingly and intentionally. AI can therefore be challenged under current constitutional case law.

Of course, courts often do not reach the merits of national security claims. Indeed, courts often dismiss cases on the basis of justiciability doctrines, limitations on constitutional *Bivens* claims, affirmative defenses for government officials like qualified immunity, and evidentiary issues such as executive privilege, classification generally, and the state secrets doctrine.[78] I will briefly discuss how AI will only make overcoming these merits-barriers harder for plaintiffs, and therefore why Congress needs to legislate to create causes of action for constitutional violations by discriminatory AI, especially in the national security context.

A. FOURTH AMENDMENT CASE LAW

1. *Fourth Amendment Case Law is Not Protective Against Profiling after* Whren

In *Whren v. United States*, the Supreme Court blocked the Fourth Amendment as a road for suing law enforcement for racial profiling, limiting suits to equal protection claims. Justice Scalia opined that "the constitutional basis for objecting to intentionally discriminatory application of laws is the Equal Protection Clause, not the Fourth Amendment," and that "[s]ubjective intentions play no role in ordinary, probable-cause Fourth Amendment analysis."[79]

In *Whren* and subsequent cases, the Court held that "the subjective intent of the officer is irrelevant" for Fourth Amendment purposes,[80] and courts would not look behind an officer's pretextual reasons for searches and seizures. As the Court later stated, "a stop or search that is objectively reasonable is not vitiated by the fact that the officer's real reason for making the stop or search has nothing to do with the validating reason," even if the real reason was "racial harassment."[81]

78. *See* Dycus et al., *supra* note 60, ch. 5 (surveying cases and presenting those doctrines).
79. Whren v. United States, 517 U.S. 806, 813 (1996).
80. Florida v. Jardines, 569 U.S. 1, 10 (2013).
81. *Id.*

Whren has been criticized roundly, firstly, for allowing and even encouraging law enforcement officers to profile so long as they can supply an excuse to search or seize someone. Public defenders are well familiar with the list of reasons why, allegedly, their clients of color were pulled over ("he said my music was too loud"; stopping too quickly at the stop sign; jaywalking, etc.).[82]

Such anecdotes illustrate a second criticism of *Whren*: that any search based on racial or other profiling conflicts with the Fourth Amendment's protection against "unreasonable" searches and seizures.[83] The NYPD's targeted surveillance of mosques and Muslim student groups across New York City and New Jersey post 9/11 might be considered inherently unreasonable, but because of the *Whren* precedent, Muslim plaintiffs did not advance that Fourth Amendment argument; they had to rely instead (successfully) on First and Fourteenth Amendment claims of disparate treatment on the basis on their religion.[84]

A third criticism of *Whren* is that by funneling all litigation under the Equal Protection Clause rather than the Fourth Amendment, the Court has severely limited the potential for redress. The *Whren* Court acknowledged the unconstitutionality of profiling ("the Constitution prohibits selective enforcement of the law based on considerations such as race"),[85] but criminal justice advocates question how easy or likely it is for profiled persons to file, much less win, equal protection claims. As discussed below in Section II.B, Equal Protection and First Amendment Establishment Clause claims are especially difficult to litigate in the national security context.

2. *Why* Whren *Should Not Apply to AI-Based Profiling*

In a 2018 case, Justice Ginsburg, at least, seemed sympathetic to growing criticisms of the *Whren* doctrine; she was concerned that it "sets the balance too heavily in favor of police unaccountability to the detriment of Fourth Amendment protection."[86] Justice Ginsburg "would leave open, for reexamination in a future case, whether a police officer's reason for acting, in at least some circumstances, should factor into the Fourth Amendment inquiry."[87]

82. Author's interview with a former client; *Whren v. United States*, 5-4 POD, https://www.fivefourpod.com/episodes/whren-v-united-states/ (last visited Jan. 13, 2024).

83. *Id.*

84. *See* Plaintiffs' Brief in Opposition to Motion to Dismiss at *8 n.3, Hassan v. City of New York, No. 2:12-cv-03401, 2014 U.S. Dist. LEXIS 20887 (D.N.J. Feb. 20, 2014).

85. *Whren*, 517 U.S. at 813.

86. District of Columbia v. Wesby, 138 S.Ct. 577, 594 (2018) (Ginsburg, J., concurring).

87. *Id.* at 594.

Perhaps that future case is here, but instead of re-examining a police officer's reason, courts will examine the programming and calculations of a government AI application. What effect AI will have on the courts' application of *Whren* is uncertain—*Whren* was, after all, written with human officers and not technology in mind. We might break down the potential scenarios for AI profiling into two cases:

(1) Where the AI is biased and/or profiling but its design also suggests a potential non-discriminatory basis for a search or seizure. Under this scenario, which is the more likely, I argue that *Whren* should not apply, and the AI[88] should be reviewable under the Fourth Amendment, because AI algorithms, data inputs and outputs, and performance are all objectively measurable and discoverable.

(2) Where the AI is purely profiling, without more. Here I argue that pre-AI case law, including *Whren*, suggests that where only profiling and no pretextual reason exists, *Whren* does not apply, and the Fourth Amendment prohibits the search or seizure.

a) When AI is profiling but provides a pretextual non-discriminatory reason, *Whren* should not apply.

Whren potentially insulates the government from Fourth Amendment claims for unreasonable searches and profiling via AI. It might do so both for intentionally biased AI and unintentionally biased AI. Given some (but not all) AI models' lack of transparency—at least at present[89]—the government might argue that courts have neither the expertise nor means to second-guess an algorithm's output: if the court cannot look into the subjective intent of a human who might be cross-examined before a jury, how can it delve into the black box of an algorithm? (Importantly, that argument is undercut by the fact that more transparent AI models are both available and being developed.)[90]

88. To include, but not limited to, all of the AI's training, testing, and validation data, inputs, algorithm, performance measurements, outputs.

89. Courts, policymakers, litigants, and legal advisors should always question whether a more transparent AI mechanism is available; increasing research into this area suggests it may be. *AI Fundamental Research – Explainability*, NAT'L INST. OF STANDARDS AND TECH. (June 16, 2022), https://www.nist.gov/artificial-intelligence/ai-fundamental-research-explainability; *see also* Brandon L. Garrett & Cynthia Rudin, *The Right to a Glass Box: Rethinking the Use of Artificial Intelligence in Criminal Justice*, 109 CORNELL L. REV. 561 (2023) (discussing the civil rights and civil liberties benefits of "Glass Box" AI models) ("As interpretable and explainable AI approaches have become more common, as subject of computer science scholarship as well as used in society, it is increasingly understood that there is a glass box alternative.").

90. *See AI Fundamental Research – Explainability supra* note 89; *see also* Garrett *supra* note 89.

Criminal defendants and civil rights plaintiffs should challenge any such assertion fiercely. Civil rights advocates should argue for access to and discovery of AI algorithms, training data, inputs, parameters, etc. One of the major (though hotly contested) claims for using AI is that it is more objective than human officers. That claim is dubious—as described above, all AI is biased and incorporates the biases of its programmers, data, and users. Some scholars have demonstrated that there is no possible way to construct an algorithm that is objectively fair to all; some criteria or values must be prioritized.[91] Nonetheless, proponents of using AI for law enforcement purposes should not have their cake—the claim of objectivity—and eat it too. One need not concede that AI is objective to insist that if the law treats it that way in allowing its use in investigations and prosecutions, then the law should treat AI likewise when citizens challenge its use. If we are to take AI proponents' claim of objectivity seriously, it follows that litigants can probe an AI's workings and those workings should be considered objectively measurable, obviating *Whren*'s concern about delving into the subjective mindset of the police officer. While some "black box" calculations may remain unknowable, plenty remains for litigants and courts to examine: the training, testing, and validation data; the algorithmic code; the inputs; the outputs; and especially any tests, which a responsible government should be performing, to measure disparate results across groups before and while deploying an AI model in real life.

Justice Scalia later wrote in *Ashcroft v. Kidd* that "the Fourth Amendment regulates conduct rather than thoughts."[92] An AI does not think, it calculates. Unlike the officer's thoughts, the AI's operating algorithm, data inputs, outputs, and test data run and in-use performance measurements, are all measurable and discoverable. A court can review them; it can review an AI's actions just as it would a police officer's conduct. Additionally, how and whether the government chooses to employ and rely on a given AI model is also human "conduct" subject to review.

---

91. *See, e.g.*, Sandra Mayson, *Bias In, Bias Out*, 128 YALE L.J. 2218, 2233 (2019).

92. Ashcroft v. Al-Kidd, 563 U.S. 731, 736 (2011) ("Fourth Amendment reasonableness 'is predominantly an objective inquiry.' . . . We ask whether 'the circumstances, viewed objectively, justify [the challenged] action.' If so, that action was reasonable 'whatever the subjective intent' motivating the relevant officials. This approach recognizes that the Fourth Amendment regulates conduct rather than thoughts; and it promotes evenhanded, uniform enforcement of the law.") (internal citations omitted); *see also* Bond v. United States, 529 U.S. 334, 338 n.2 (2000) ("The parties properly agree that the subjective intent of the law enforcement officer is irrelevant in determining whether that officer's actions violate the Fourth Amendment . . . . This principle applies to the agent's acts in this case as well; the issue is not his state of mind, but the objective effect of his actions.").

Civil rights litigants might thus argue that *Whren* is inapplicable to the extent that the government relies on the objectively measurable calculations of an AI rather than the subjective, intuitive reasoning of a human being.[93] *Whren* was not written with AI in mind. Moreover, the government should be consistently measuring the AI's performance and biases before and during its operation, so the government should know, objectively and at all relevant times, what those measurements are. The objectively measurable nature of AI is demonstrated by the fact that all of its crucial elements are expressed by AI in numbers: those elements include the AI's training, testing, and validation data; inputs used in the field; algorithmic formula(e); and outputs.

If an AI is alleged or shown to be biased—to rely on suspect data or inputs or proxies for suspect data or inputs or to otherwise produce disparate results—then *at a minimum*, a court should consider whether the AI also used non-suspect factors and inputs to help reach its conclusions, and if so, whether those *independently* would establish probable cause (or the Fourth Amendment predicate at issue). *Whren* itself requires no less. *Whren* dictates that the government's behavior must be "objectively justifiable,"[94] and indeed the Court found a basis for probable cause, independent of racism, on the facts in that case.[95] The government should have to demonstrate an objective basis for the AI's outputs—and that will require transparency, at least with some reviewing body, perhaps the court *in camera*, if not with the litigants themselves (which due process may well require).

But even this minimum treatment seems insufficient and problematic. It could mask the fact that but for the discriminatory bias in the AI, the individual defendant or investigatory target would not have been flagged for further scrutiny. The fact remains (it was never conceded above) that AI is not objective or trustworthy, or in many cases, transparent. And the mathematically inputted biases may have infected the objectivity of any and all calculations. Litigants should certainly argue for throwing out AI results altogether, if the AI is shown to be infected with bias. If so infected, it will not

93. This argument may be bolstered by Professor Andrew Guthrie Ferguson's suggestion that courts might be less deferential to law enforcement's use of facial recognition systems: "First, much of the Supreme Court's expansion of police power can be traced to deference to human decision-making, and when decision-making is made at a programmatic or administrative level, such deference wanes . . . . Second, while the Supreme Court seems to forgive isolated errors or pretextual biases of individual officers, the Court does not forgive recurring errors or systemically biased decisions." Andrew G. Ferguson, *Facial Recognition and the Fourth Amendment*, 105 MINN. L. REV. 1105, 1191 (2021).

94. *See Whren*, 517 U.S. at 812.

95. *Id.* at 819.

provide the independent, objective basis for probable cause that *Whren* requires.[96]

Where we can objectively measure AI bias, violations of the reasonableness and probable cause requirements of the Fourth Amendment are demonstrable and evident.

b) When AI is only profiling, *Whren* does not apply, and the discriminatory AI cannot contribute to probable cause.

Civil rights advocates might also argue that where AI is openly and only profiling—that is, where it was programmed to profile or is being used so obviously to profile as to preclude other sources of probable cause for search or seizure—it falls outside the *Whren* analysis, even under pre-AI case law. (Being "programmed to profile" might include, among other things, using suspect categories, but also proxies for suspect categories.) *Whren* spoke to cases where there was at least a pretextual objective basis for probable cause—that is, where law enforcement claims to search or arrest someone for reasons other than, and independent of, racial profiling. When law enforcement explicitly or obviously profiles, and there is no other sufficient basis for the government action, does *Whren* still block a Fourth Amendment claim? At least one district court addressed that question in *Farag v. United States*,[97] and answered No: Where no legitimate reason for seizure existed, *Whren* did not apply. It determined that the *Whren* Court established only that "an officer's subjective, potentially race-based motivations were irrelevant to the Fourth

96. A potential critique of the argument advanced here, that review of AI under the Fourth Amendment should not be precluded by *Whren*, is that *Whren*'s primary concern was not the difficulty of proving that the police officer's subjective intent was racist, but rather whether objective, reasonable probable cause existed. Justice Scalia wrote that the precedent cases he examined in W*hren* were not based only upon the evidentiary difficulty of establishing subjective intent; rather, "their principal basis" was that "the Fourth Amendment's concern with 'reasonableness' allows certain actions to be taken in certain circumstances, whatever the subjective intent." *Whren*, 517 U.S at 814.

The evidentiary issue of measuring subjective intent was, however, a motivating concern. As referenced above, years later, Justice Scalia wrote that conduct, rather than thoughts, is reviewable under the Fourth Amendment. If the Fourth Amendment's concern is conduct, biased AI is a bad actor. Moreover, if "the Fourth Amendment's concern with 'reasonableness' allows certain actions to be taken in certain circumstances," then the government must prove that its AI enables those objectively reasonable actions. *Id.* And therefore, the AI, its workings, and calculations, become reviewable under the Fourth Amendment, as argued above. The government must prove that its AI is not making recommendations or predictions based in part or in whole on suspect categories or proxies for those categories. The government will need to show that it had legitimate reasons for a search or seizure, reasons independent of and *uninfluenced by* algorithmic bias, which may prove very difficult to do.

97. Farag v. U.S., 587 F. Supp. 2d 436, 461–65 (E.D.N.Y. 2008).

Amendment *once probable cause is established*; the *Whren* Court was not called upon to address whether race might be relevant to the probable-cause analysis itself."[98]

The *Farag* court then rejected the government's open assertion that the plaintiffs' ethnicity could be a factor in determining the validity of their seizure and detention.[99] Surveying federal case law, the court divided the use of "race in the context of the Fourth Amendment" into three categories:

(1) The "least controversial" use of 'race' as an "identifying factor," where a victim or witness of a crime describes the perpetrator using racial description, among other physical descriptors, such as what the individual was wearing.[100]

The *Farag* court opined that "there can be little doubt that law enforcement officials may consider that description in deciding whom to detain, even though the description is based, in part, on race."[101] I would suggest, however, that even this use of 'race' might be problematic, for example, if law enforcement officers detain a large number of innocent people on the basis of 'race.' AI in particular risks that scenario: if a racial, ethnic, or religious "identifying factor" is used as an input to query a database, the AI might identify and output a large number of innocent people who then come under additional scrutiny on that racial or other basis.

(2) The "so-called 'racial incongruity' argument—i.e., that race is indicative of criminality when members of a particular race seem 'out of place' in a particular location."[102]

This is the type of profiling infamously associated with the unwarranted arrest of Professor Henry Louise Gates, Jr., outside his own home in Cambridge, Massachusetts. Surveying case law, the *Farag* court found that some courts had "sidestepped the issue by finding probable cause or reasonable suspicion based on some other, non-racial factors," but those courts that had "squarely

98. *Id.* at 462. That conclusion is consistent with *Jardines*, 569 U.S. at 10 (2013), where racial profiling was not at issue, but like *Whren* was authored by Justice Scalia ("The State points to our decisions holding that the subjective intent of the officer is irrelevant. *See* Ashcroft v. al-Kidd, 563 U.S. 731 (2011); Whren v. United States, 517 U.S. 806 (1996). But those cases merely hold that a stop or search *that is objectively reasonable* is not vitiated by the fact that the officer's real reason for making the stop or search has nothing to do with the validating reason. Thus, the defendant will not be heard to complain that although he was speeding the officer's real reason for the stop was racial harassment. *See id.*, at 810, 813. Here, however, the question before the court is precisely *whether* the officer's conduct was an objectively reasonable search." [The Court concluded it was not.]).

99. *Farag*, 587 F. Supp. 2d at 443.

100. *Id.* at 462.

101. *Id.*

102. *Id.*

addressed the incongruity argument ha[d] uniformly rejected it." Any AI profiling on the basis of "incongruity" should likewise be rejected as a basis for a Fourth Amendment predicate.

(3) "Propensity" profiling, at issue in *Farag*, where the government made the bold-faced argument "that plaintiffs' Arab ethnicity is a relevant consideration [in a probable cause determination] is premised on the notion that Arabs have a greater *propensity* than non-Arabs toward criminal activity—namely, terrorism."[103]

The *Farag* court rejected any use of "propensity" profiling as the basis for finding probable cause or contributing[104] to a finding of probable cause. "Even granting that all of the participants in the 9/11 attacks were Arabs . . . the likelihood that *any given airline passenger* of Arab ethnicity is a terrorist is so negligible that Arab ethnicity has no probative value in a particularized reasonable-suspicion or probable-cause determination."[105] The court concluded that the bulk of precedent "clearly evidences . . . an increasing hostility to the use of race as a basis for police action under the Fourth Amendment," and that the "specter of 9/11" did not provide justification for propensity profiling.[106]

If an AI uses a suspect category or a proxy for a suspect category to predict criminal or terrorist behavior, that is tantamount to propensity profiling. Likewise, no matter how discriminatory bias enters an AI model, if the biased AI is used to predict illegal behavior, that is likewise propensity profiling. And whether the biased AI outputs inform the government's probable cause showing in whole or in part, that showing will be invalid under the Fourth Amendment. In other words, race or ethnicity-based factors (biased AI) may not be used to establish criminal propensity under the Fourth Amendment.

---

103. *Id.* at 463. The government relied on what the court referred to as "dictum" from a 1975 Supreme Court case, *United States v. Brignoni-Ponce,* that "'the likelihood that any given person of Mexican ancestry is an alien is high enough to make Mexican appearance *a relevant factor*' in the Fourth Amendment calculus, if it were not *the only* basis for suspicion.'" *Farag*, 587 F. Supp. 2d at 465 (quoting *United States v. Brignoni-Ponce*, 422 U.S. 873, 886–87 (1975)). The *Farag* court could find no case, however, that had "ever marshaled statistics to conclude that racial or ethnic appearance is correlated with, and thus probative of, any type of criminal conduct *other than* immigration violations," and noted that the Ninth Circuit, where *Brignoni-Prince* originated, had found twenty-five years later "that the statistical inference on which it was based was no longer valid, even in its original illegal-immigration context." *Farag*, 587 F. Supp. 2d at 464.

104. *Farag*, 587 F. Supp. 2d at 466–67.

105. *Id.* at 464. This is, notably, the opposite of the logic adopted by Justice Kennedy (for First and Fifth Amendment purposes) in *Iqbal*, discussed below.

106. *Farag*, 587 F. Supp. 2d at 467 (internal quotations omitted).

As suggested above, AI also blurs the line between using 'race' or ethnicity as an "identifying factor" to describe a known criminal suspect and using those categories in propensity profiling. If, on the basis of a suspect description entered into an AI model, the government surveils or detains tens or hundreds or thousands of individuals on the basis of 'race' or ethnicity, that begins to look more like propensity profiling in effect, even if not (necessarily) in motivation. To analogize to the reasoning in *Farag*, the larger the number of people the AI screens or suggests, the less statistically relevant the original suspect description becomes, because the likelihood that *any given* individual screened is the actual suspect becomes negligible. The technology collapses the categories (identifying factor and propensity); it may render them no longer distinguishable or useful.

Farag and its discussion of precedent suggest a clear avenue for arguing that *Whren* should not preclude a court from invalidating a probable cause or similar determination made in whole or in part on AI's use of racial, ethnic, or other similar factors.

3. Other Bases to Challenge AI Surveillance Under the Fourth Amendment

Whren and profiling issues aside, there is some reason to hope that the Court will be skeptical of the reasonableness of AI-based searches in the context of the Fourth Amendment. In *Carpenter*, Chief Justice Roberts determined that the use of another, newish technology—over 127 days' worth of cell site location information used to retroactively pinpoint the defendant's whereabouts during that period—required a warrant to be reasonable. Likewise, future courts may insist on individual warrants presented to neutral magistrates before AI outputs are used to establish the predicate for a search or seizure. Where AI is used to process information from sensors such as video cameras or to power autonomous drones or other surveillance with video and sound sensors, a court might be even more worried about "near perfect surveillance,"[107] and require warrants, as some states have begun to do.[108] The First Amendment chilling effects of such surveillance are well established in scholarly literature.[109]

But even *Carpenter* notably carved national security investigations out of its holding, declining to opine on them or require a warrant. The Court cautioned that it did "not disturb" the application of the Fourth Amendment third-party

107. *Carpenter*, 585 U.S. at 312.

108. *See Drone Laws by State*, FINDLAW (July 12, 2021), https://www.findlaw.com/consumer/consumer-transactions/drone-laws-by-state.html.

109. *See, e.g.*, Kaminski, *supra* note 67.

doctrine, which is critical to government surveillance.[110] Nor did it "call into question conventional surveillance techniques and tools, such as security cameras," nor "consider other collection techniques involving foreign affairs or national security."[111]

Even a court skeptical of AI might be more inclined to show deference to the executive with respect to a Fourth Amendment issue affecting foreign affairs or national security.[112]

In sum, courts might decline to apply *Whren* to where the reasonableness of AI-based searches and seizures is in doubt. They might also apply warrant requirements to more invasive AI-based technologies, as suggested by past cases involving new technologies, such as *Carpenter*, as well as *Kylo v. United States* (requiring a warrant for the search of house interior via infrared technology), and *Riley v. California* (requiring a warrant for the search of cell phone contents).[113] *Carpenter*'s carveouts for video surveillance and national security, however, might temper any such hope by privacy advocates. And, in the absence of warrant requirements, if courts do apply *Whren* against litigants who challenge the government's use of biased AI to establish reasonable suspicion or probable cause, the narrow avenue for any redress will be the Equal Protection Clause.

B. EQUAL PROTECTION AND FIRST AMENDMENT RELIGION CLAIMS

National security-related equal protection claims seeking injunctions of government policies have had only limited success, while *Bivens* claims for monetary damages after-the-fact have faced closed courthouse doors. This section addresses injunctive relief. The next section, III.C, addresses the non-violability of national security suits for *Bivens* damages for any constitutional harms.

Though there is a dearth of cases that have ever reached the merits for injunctive relief, regardless of whether the courts apply heightened scrutiny or

110. *Carpenter*, 585 U.S. at 316.

111. *Id.*

112. *See, e.g.*, *In re* Directives to Yahoo! Inc. Pursuant to Section 105B of Foreign Intelligence Surveillance Act, 551 F.3d 1004 (FISA Ct. Rev. 2008); *c.f.*, United States v. U.S. District Court (Keith), 407 U.S. 297, 308 (1972) (requiring a warrant for wiretapping in 'domestic' national security investigations but explicitly not addressing whether a warrant would be required in an investigation relating to a foreign power: "the instant case requires no judgment on the scope of the President's surveillance power with respect to the activities of foreign powers, within or without this country.").

113. *See* Fourth Amendment discussion in James Baker, Laurie Hobart & Matthew Mittelsteadt, *An Introduction to Artificial Intelligence for Federal Judges*, FED. JUD. CTR. (Feb. 13, 2023), https://perma.cc/7DJT-T9D2; *see also Kyllo v. United States*, 533 U.S. 27 (2001); *Riley v. California*, 573 U.S. 373 (2014).

rational basis review, the government's interest in national security arguments often (but not always) trumps. This section examines the significant barriers to injunctive relief for victims of discriminatory AI bias.

But I also argue why injunctive relief may be plausible, even required, under current equal protection precedent, if the courts were to reach the merits of claims. Using biased AI does more than cause disparate impact; it is intentional discrimination. Any government use of biased AI constitutes a facially discriminatory policy; or in the alternative, the government's knowing choice to use biased AI demonstrates purposeful disparate treatment.

1. Barriers to Injunctive Relief

While essential, injunctive relief for equal protection and First Amendment violations caused by discriminatory AI is unlikely, especially in national security contexts. There are many barriers to injunctive relief. Some barriers are specific to AI, and all might be triggered and exacerbated by the hidden use and pervasive nature of AI. I outline five here:

(1) discrimination suits must allege intentional disparate treatment, not just disparate impact;

(2) where courts do find disparate treatment and apply strict scrutiny, the government's national security interest will always be significant;

(3) after *Trump v. Hawaii*,[114] courts may apply rational basis review rather than strict scrutiny in certain cases—and it is unclear which cases;

(4) courts may find 'affirmative action'-like interventions to rid AI of bias problematic after *Students for Fair Admissions, Inc. (SFFA) v. President & Fellows of Harvard College*;[115] and

(5) injunctive relief may be impracticable or impossible for victims of profiling to seek at the time of injury.

Throughout this discussion, I also offer arguments for why, even under current case law, courts should still provide relief to litigants alleging harm from discriminatory AI.

a) Plaintiffs must allege intentional disparate treatment, not just disparate impact.

One of the first challenges plaintiffs may face is establishing that they are victims of disparate treatment, not just disparate impact. Only disparate

114. *Trump v. Hawaii*, 585 U.S. 667 (2018).

115. Students for Fair Admissions, Inc. v. President & Fellows of Harvard College, 600 U.S. 181 (2023).

treatment is actionable under equal protection law precedent:[116] "To state an equal-protection claim, Plaintiffs must allege (and ultimately prove) 'intentional discrimination.'"[117] In 1976, the Supreme Court stated in *Washington v. Davis* that "our cases have not embraced the proposition that a law or other official act, without regard to whether it reflects a racially discriminatory purpose, is unconstitutional solely because it has a racially disproportionate impact."[118] In the 2009 case *Ashcroft v. Iqbal*, Justice Kennedy wrote for the Court that in both First and Fifth Amendment discrimination cases, a "plaintiff must plead and prove that the defendant acted with discriminatory purpose."[119]

The standard for pleading such "purposeful discrimination" is formidable, requiring "more than 'intent as volition or intent as awareness of consequences."[120] Rather, as Justice Kennedy wrote in *Ashcroft v. Iqbal*, "purposeful discrimination" "involves a decisionmaker's undertaking a course of action 'because of, not merely in spite of,' the action's adverse effects upon an identifiable group."[121] Plaintiff Iqbal would have had to plead sufficient facts to show that government defendants "adopted and implemented the detention policies at issue not for a neutral, investigative reason but for the purpose of discriminating on account of race, religion, or national origin."[122]

Iqbal alleged that on account of his race, religion, and/or national origin, the FBI designated him a person of "high interest" and detained him as part of its 9/11 investigation. He further alleged defendants John Ashcroft and Robert Mueller had approved "[t]he policy of holding post-September-11th detainees in highly restrictive conditions of confinement" and "knew of, condoned, and willfully and maliciously agreed to subject" Iqbal to harsh conditions of confinement solely on the basis of "his religion, race, and/or

116. This "anticlassification" approach has been criticized by many. *See, e.g.*, Charles Lawrence, *supra* note 38, at 322 (". . . [R]equiring proof of conscious or intentional motivation as a prerequisite to constitutional recognition of that decision is race-dependent ignores much of what we understand about how the human mind works. It also disregards both the irrationality of racism and the profound effect that the history of American race relations has had on the individual and collective unconscious.").

117. Hassan v. City of New York, 804 F.3d 277, 294–95 (2015) (citing Washington v. Davis, 426 U.S. 229, 241 (1976) and Pers. Adm'r of Mass. v. Feeney, 442 U.S. 256, 276 (1979)).

118. Washington v. Davis, 426 U.S. 229, 239 (1976).

119. Ashcroft v. Iqbal, 556 U.S. 662, 676 (2009) (citing Church of Lukumi Babalu Aye, Inc. v. Hialeah, 508 U.S. 520, 540–541 (1993) (opinion of Kennedy, J.) (First Amendment); Washington v. Davis, 426 U.S. 229, 240 (1976) (Fifth Amendment)).

120. *Id.*

121. *Id.* at 676–77 (internal citations omitted).

122. *Id.*

national origin."[123] He alleged that Ashcroft was the "principal architect" of the policy and Mueller was "instrumental in [its] adoption, promulgation, and implementation."[124]

The Court found disparate treatment had not been sufficiently pleaded. The Court first observed that the 9/11 attackers were "[nineteen] Arab Muslim hijackers who counted themselves members in good standing of al Qaeda, an Islamic fundamentalist group."[125] From that starting point, the Court then concluded that it was not surprising that "a legitimate policy directing law enforcement to arrest and detain individuals because of their suspected link to the attacks would produce a *disparate, incidental impact* on Arab Muslims, even though the purpose of the policy was to target neither Arabs nor Muslims."[126]

But it *is* problematic that a search for religious *fundamentalists* in Al Qaeda resulted in the detention of "Arab Muslims." The leap from detaining Al Qaeda fundamentalists to detaining people based on their Arab origin or ethnicity or their membership in world's second largest religion describes something far from "incidental" impact or causation. The opinion makes clear that a "disparate, incidental" impact will not be enough to show purposeful discrimination,[127] even where even where the leap from "terrorist" to "Arab Muslim" is long indeed and itself demonstrative of profiling. The equivalent would be to detain American white males of the Oklahoma City bomber Timothy McVeigh's religious or cultural background, before he was radicalized, and contend there was no intentional discrimination, just incidental impact.

For AI, this standard suggests that it may not be enough for plaintiffs to allege that the government relied on an algorithm to make investigatory and arrest decisions with disparate impacts on certain populations. An easy parallel to Justice Kennedy's line of reasoning would be for the government to argue that correlations between algorithmic outputs and suspect classes are not purposeful, but rather, incidental.

I write that it "may" not be enough for plaintiffs to point out the disparate results of algorithms, rather than it "will" not be enough. With proper diligence the government should be on notice of the likelihood of such "incidental" impacts, such that any choice to go forward with such an algorithm will be, as I argue below, quite purposeful; moreover, the algorithm itself might be viewed as facially discriminatory. But one can imagine the government urging

123. *Id.*
124. *Id.*
125. *Id.* at 682.
126. *Id.* (emphasis added).
127. *Id.*

the reprehensible argument that because an algorithm is trained on a limited set of input data—the nineteen 9/11 hijackers, for extreme example—we should uncritically accept that its predictive outputs may disproportionately reflect suspect category features, such as "Arab" or "Muslim" or "male," even though those are not the causal or determinative feature—al Qaeda membership—in predicting terrorist activity. The problem is that the results will be both under- and over-inclusive. The results will be dangerously underinclusive because there are many potential terrorists, including white supremacist domestic terrorists, as is well documented in national security law scholarship,[128] who would not fit the profile. The results will be harmfully, grossly overinclusive by profiling swaths of innocent Arabs and Muslims and adding them unnecessarily to the cast of the investigative net.

i) Why using biased AI constitutes disparate treatment

There is room for hope—and a strong argument that discriminatory AI bias creates disparate treatment, not just disparate impact. I suggest two potential theories to show purposeful discrimination:

(1) an algorithm itself is facially discriminatory when it yields biased results, and;

(2) the government, which is on notice that AI is potentially discriminatory, acts purposefully and intentionally when it adopts a discriminatory AI model.

In pre-AI contexts, at least two courts have wrestled with the fine distinction between disparate impact and disparate treatment and still held in favor of plaintiffs, finding disparate treatment as required under the *Washington v. Davis* standard.[129] In *Hassan v. City of New York*, plaintiffs alleged that the New York Police Department (NYPD) conducted a wide-scale, secret surveillance program of Muslims in their businesses, houses of worship, organizations, and schools in New York City, New Jersey, and other surrounding states.[130] In *Floyd v. City of New York*, plaintiffs alleged that NYPD's use of stop and frisk violated their Fourth Amendment and equal protection rights, where "[o]ver 80% of [NYPD's] 4.4 million stops [between January 2004 and 2012] were of blacks and Hispanics."[131]

In *Hassan*, the district court stated it was not enough for plaintiffs to allege that they were "Muslim and that the NYPD surveilled more Muslims than

128. *See, e.g.*, Sinnar, *supra* note 44, at 1388–92.

129. As noted above, *Washington v. Davis*, 426 U.S. 229 (1976), established that laws that have a disparate impact but were not adopted to advance a discriminatory purpose do not violate the Equal Protection Clause.

130. Hassan v. City of New York, 804 F.3d 277, 285 (3d. Cir. 2015).

131. Floyd v. City of New York, 959 F. Supp. 2d 540, 540 (S.D.N.Y. 2013).

members of any other religion. Rather, Plaintiffs' religious affiliation must have been a substantial factor in that different treatment."[132] The court suggested there were a "variety" of ways for plaintiffs in equal protection suits to prove disparate treatment. Plaintiffs could:

1. "point to a policy that is facially discriminatory, meaning that the policy by its own terms singles out Muslims for different treatment,"

2. "identify a policy that either shows no classification on its face or else indicates a classification which seems to be legitimate, yet one that NYPD officers apply to Muslims with a greater degree of severity than other religious groups," or

3. "identify a facially neutral policy that the City purposefully designed to impose different burdens on Muslims and that (even if applied evenhandedly) does in fact."[133]

The *Hassan* court accepted plaintiffs' allegations as plausible and sufficient to survive a motion to dismiss under the first theory: a facially discriminatory policy. Plaintiffs alleged specifics about the surveillance program: when it was conceived; where the City implemented it; why it had been employed ("because of a the belief that Muslim religious identity . . . is a permissible proxy for criminality,"); and how (via a "variety of methods" including videos, photographs of mosques, businesses, schools, monitoring of websites, listservs and social media, and use of undercover officers to monitor neighborhoods and mosques).[134] Notably, the court stated, "[t]hat we might conjure up some non-discriminatory motive to explain the City's alleged conduct is not a valid basis for dismissal."

---

132. *Hassan*, 804 F.3d at 294–95 (internal citations omitted).

133. *Id.* (internal quotations omitted); *see also Floyd*, 959 F. Supp. 2d at 660–61 ("Racial profiling constitutes intentional discrimination in violation of the Equal Protection Clause if it involves any of the following: an express classification based on race that does not survive strict scrutiny; the application of facially neutral criminal laws or law enforcement policies 'in an intentionally discriminatory manner;' or a facially neutral policy that has an adverse effect and was motivated by discriminatory animus."); Brown v. City of Oneonta, 221 F.3d 329, 337 (2d Cir. 1999) ("There are several ways for a plaintiff to plead intentional discrimination that violates the Equal Protection Clause. A plaintiff could point to a law or policy that "expressly classifies persons on the basis of race." *Id.* (citing Adarand Constructors, Inc. v. Pena, 515 U.S. 200, 213, 227–29 (1995)). Or, a plaintiff could identify a facially neutral law or policy that has been applied in an intentionally discriminatory manner. *See* Yick Wo v. Hopkins, 118 U.S. 356, 373–74 (1886). A plaintiff could also allege that a facially neutral statute or policy has an adverse effect and that it was motivated by discriminatory animus. *See* Village of Arlington Heights v. Metropolitan Hous. Dev. Corp., 429 U.S. 252, 264–65 (1977); Johnson v. Wing, 178 F.3d 611, 615 (2d Cir. 1999).").

134. *Hassan*, 804 F.3d at 295.

Civil rights litigants might model claims for AI-profiling accordingly. Plaintiffs alleging harms from AI surveillance, even without a copy of the policy or access to the algorithm that harmed them, might be able to pinpoint when surveillance began—at least when plaintiffs started feeling the effects of it; where it appeared to happen (for instance, at airports, with facial recognition cameras); why (because of certain overlapping identity traits they seemed to have in common, suggesting that the government considered such traits a proxy for criminal, terrorist, or counterintelligence activity); and how (via airport or entry point screening, or surveillance cameras). Such a showing might add up to establish a facially discriminatory algorithm—i.e., one that takes into account (or fails to take into account) suspect categories such as 'race' or religion or uses proxies for those categories, such as zip code or social organization membership.

However, it may be difficult if the profiling is less AI obvious, both because of the surreptitious nature of the alleged technological surveillance and any classification of the programs. One thinks of cases like C*lapper v. Amnesty International*,[135] where plaintiffs lacked standing because they could not prove beyond speculation that they were subject to the surveillance under section 702 of the amended Foreign Intelligence Surveillance Act of 1978, and *Zaidan v. Trump*, where a plaintiff likewise could not show that he was on a "kill list,"[136] or if so, that it was a United States rather than foreign operated list.[137] (For this reason, additional oversight of classified AI programs used by the government is needed, ideally by a court, as recommended in Part V below.)

If it can be shown that an AI model is discriminatory, either through circumstantial evidence as in *Hassan*, or with access[138] to the actual algorithm and its outputs, then the argument that the AI itself is facially discriminatory has several advantages.

First, it strikes me as the most genuine characterization of biased AI. Regardless of what went into the programming, if the end result of the AI is to cabin suspect classes differently, then the AI and any government policy implementing its outputs are facially discriminatory. The AI program "explicitly draws racial lines,"[139] if not explicitly in its programming, then

135. Clapper v. Amnesty Int'l USA, 568 U.S. 398, 422 (2013).

136. Zaidan v. Trump, 317 F. Supp. 3d 8, 19 (D.C. 2018).

137. *Id.* at 28.

138. After-the-fact (of alleged discrimination), access might be gained through discovery or *in camera* judicial review. An oversight court or entity might also require access to an algorithm and an assessment of its results before it is deployed, as suggested in Part V below.

139. *See* GEOFFREY R. STONE, LOUIS M. SEIDMAN, CASS R. SUNSTEIN, MARK V. TUSHNET, PAMELA S. KARLAN, AZIZ Z. HUQ & LEAH M. LITMAN, CONSTITUTIONAL LAW 483 (9th ed. 2023) (discussing facially discriminatory laws).

somewhere within its workings, such that its outputs fall within those lines. The AI is using 'race' or another category as the basis for a burden or disadvantage. If the AI's outputs separate people in ways that track racial or other classifications, the AI is facially discriminating. The AI is explicitly fencing people into different groups—it may be using a language of numbers, not words, but it is still openly fencing them, and the government is adopting that segregating system.[140]

For plaintiffs, treating biased AI as facially discriminatory has another advantage: they need not show discriminatory purpose for courts to apply strict scrutiny (for suspect categories that would correlate with strict scrutiny, such as 'race').[141] If a government agency uses an AI model that produces objectively measurable discriminatory results when tested, then the government is adopting a facially discriminatory program, even if it was never explicitly programmed with "race" or "ethnicity" or other suspect classifications in its algorithm or the labels it assigns data.

A final advantage of the facially-discriminatory-AI theory is that it fits well into the Court's "anticlassification" theory of equal protection law. The AI is classifying, and that is an equal protection violation (unless a compelling governmental reason can be shown and the government policy is narrowly tailored, etc.). Importantly, such a theory allows for data engineering or modeling to *reduce* bias. Any engineering or modeling done to reduce bias would produce (something closer to) facially neutral results. It would look less like affirmative action, which the Court has rejected, and more like an *anti*classification tool, in line with the Court's precedent and philosophy of equal protection. I will discuss this topic further below.

If a court does not accept that a biased AI model is facially discriminatory, plaintiffs can still argue that the government's choice to use the AI shows purposeful discrimination.[142] The government is on notice, as evidenced by its

140. In *Automating Judicial Discretion: How Algorithmic Risk Assessments in Pretrial Adjudications Violate Equal Protection on the Basis of Race*, 40 MINN. J. L. & INEQ. 371, 388–89 (2022), Christopher Thomas and Antonio Pontón-Núñez argue that criminal risk assessments should be treated as facially discriminatory because they explicitly use suspect classifications like race or proxies for those classifications. I agree with that analysis, and add here that even if the inputs, including proxies, cannot be discovered, the results or outputs can be evaluated for disparate treatment across groups. While we cannot ignore racist (as opposed to race-conscious) inputs and other design elements, there are advantages to emphasizing outputs given the Court's anticlassification view of equal protection law, as I argue here and below in relation to race-conscious engineering interventions against biased AI modeling.

141. *See* ERWIN CHEMERINSKY, CONSTITUTIONAL LAW, PRINCIPLES AND POLICIES 742–43 (7th ed. 2023).

142. *See* Thomas and Pontón-Núñez, *supra* note 140, at 398 ("[A] *Floyd* [discriminatory] intent framework could be applied to algorithmic [criminal risk assessments] because state

own policies, that AI often discriminates. The government is well-versed in all the ways that conscious and unconscious bias can enter AI data and design. Responsible use suggests that the government must monitor the AI for bias at all stages of the machine learning lifecycle—from design and conception through use and maintenance. If the AI discriminates, the government should know. Even if the AI was not consciously programmed to discriminate, but it nevertheless does so and the government still chooses to use it, that suggests an intentionally discriminatory application.[143]

In *Floyd*, the court held that plaintiffs showed an intentionally discriminatory application of a facially neutral policy,[144] the second theory outlined above for showing disparate treatment. Plaintiffs' "statistical evidence of racial disparities in stops [was] sufficient to show a discriminatory effect."[145] The *Floyd* plaintiffs showed that, always controlling for other relevant variables: the NYPD carried out more stops where there were more Black and Hispanic residents; NYPD officers were more likely to stop Black and Hispanic people than white people *within* precincts; and NYPD officers were more likely to use force against Black and Hispanic people than white people.[146] The *Floyd* plaintiffs also showed that NYPD officers stopped Black and Hispanic people with less justification than white people.[147] This "statistical evidence of a racially disproportionate impact" was supplemented with significant anecdotal evidence.[148]

Floyd may suggest a way forward for AI civil rights litigants. The fine line between non-actionable "disparate impact" and actionable "statistical evidence of racial disparities . . . sufficient to show a discriminatory effect"[149] rising to the level of disparate treatment, seems to be that officials intentionally apply the facially neutral policy in a discriminatory manner. In *Floyd*, that intent was shown by the fact that when, *holding other factors constant*, the police's actions still

---

actors have both deliberately ignored the adverse effects on Black and Latino defendants, as well as mandated their use without consideration of scientific studies warning against their use.").

143. Depending on the facts, plaintiffs could also argue the third theory outlined in *Hassan*: "identify a facially neutral policy that the City purposefully designed to impose different burdens' on Muslims and that (even if applied evenhandedly) does in fact."*See Hassan*, 804 F.3d at 294.

144. *Floyd*, 959 F. Supp. 2d at 661. They also showed that the City had violated the Equal Protection Clause under the first method of proof, an "express classification based on race that does not survive strict scrutiny," "insofar as the use of race [was] explicit.".

145. *Id.* at 661.

146. *Id.*

147. *Id.*

148. *Id.* at 661–62.

149. *Id.* at 661.

had a statistically demonstrable discriminatory effect, as well as by anecdotal evidence suggesting intent, such as differences in how officers reported stops. In national security cases, anecdotal evidence may not be readily available absent leaks, but litigants may be able to show that, holding other factors constant, they have been treated disparately. For example, information about who is stopped for further questioning at airports is available, as is information about who was detained in maximum security prisons post-9/11.

Perhaps most helpful for AI plaintiffs, the *Hassan* court emphasized that a claim of disparate treatment requires proving intent but not invidious motive. It rejected the City's argument that even if plaintiffs had alleged a facial classification based on religious affiliation, their suit should be dismissed if the more likely explanation for NYPD's actions was public safety rather than religious discrimination.[150] The court distinguished between intent, which "asks 'whether a person acts intentionally or accidentally,'" and motive, which "asks 'if he did it intentionally, why did he do it?'"[151] "Invidious" motive, the *Hassan* court wrote, is not necessary for discriminatory intent. "All you need is that the state actor *meant* to single out a plaintiff because of the *protected characteristic* itself."[152] Citing *Floyd*, among other cases, for the proposition that intentional discrimination need not be motivated by ill will, enmity or hostility to contravene the Equal Protection Clause, the court concluded, "[t]hus, even if NYPD officers were subjectively motivated by a legitimate law-enforcement purpose (no matter how sincere), they've intentionally discriminated if they wouldn't have surveilled Plaintiffs had they not been Muslims."[153] Likewise in *Floyd*, to demonstrate discriminatory intent, plaintiffs had to show that "those responsible for the profiling did so at least in part because of, not merely in spite of its adverse effects upon the profiled racial group," but not that "race was the sole, predominant, or determinative factor in a police enforcement action," nor that the discrimination was "based on ill will, enmity, or hostility."[154]

Civil rights advocates might show that the AI or its use was intentionally discriminatory, even if not invidiously motivated. As discussed above, AI inputs, outputs, algorithms, weights, data, labels, and performance results are all knowable (and discoverable, if courts so determine). If AI is programmed with suspect categories or proxies, that looks quite intentional. The

150. *Hassan*, 804 F. 3d at 297.

151. *Id.* at 297 (quoting John William Slamon, JURISPRUDENCE § 134 at 398 (7th ed. 1924)).

152. *Id.* at 297.

153. *Id.* at 298.

154. *Floyd*, 959 F. Supp. 2d at 661 (internal quotations omitted).

government, on notice that AI often discriminates, has a duty to monitor the AI for bias at all stages of the machine learning lifecycle. If the AI is or becomes discriminatory, the government should know. And if the government knows and uses the AI anyway, that is or at least strongly suggests intentional disparate treatment.

b) The government will always have a compelling interest in national security cases.

Even if a plaintiff successfully pleads facial discrimination or disparate treatment[155] by a government AI program, the program may still be upheld if it survives strict scrutiny. Chief Justice Roberts recently described strict scrutiny as a "daunting two-step examination," where a court asks "first, whether the racial classification is used to 'further compelling governmental interests'" and "[s]econd, if so, . . . whether the government's use of race is 'narrowly tailored'—meaning 'necessary'—to achieve that interest."[156]

How the Court might apply that standard to national security-related AI cases is unknown. On the one hand, in national security cases, the government interest will almost always be deemed compelling. And courts have a history of often, though not always, deferring to the executive branch in matters of national security.[157] Even strict scrutiny leaves considerable room for judicial discretion; personal ideologies of judges may determine whether an outcome looks more like *Korematsu* or *Hassan v. City of New York*.

155. GEOFFREY R. STONE, LOUIS M. SEIDMAN, CASS R. SUNSTEIN, MARK V. TUSHNET, PAMELA S. KARLAN, AZIZ Z. HUQ & LEAH M. LITMAN, CONSTITUTIONAL LAW 483 (9th ed. 2023) ("After Washington v. Davis, a court confronted with a classification that disadvantages a racial minority must first determine whether it constitutes a 'racial classification.' If it does—either because the classification explicitly draws racial lines or because it is motivated by a racial purpose—the court will use strict scrutiny and probably invalidate it."). *Cf.* CHEMERINSKY, *supra* note 141, at 776 ("If a law is racially neutral a challenger must show a discriminatory purpose and a discriminatory effect. If such proof is provided, the government has the opportunity to demonstrate that it would have taken the same action regardless of race or national origin. If the Court accepts the government's justification and rejects the claim of a discriminatory purpose, only rational basis review is used. If the Court is convinced that there is a discriminatory purpose, the law is treated as a race or national origin classification and the law will be invalidated. The formal application of strict scrutiny is unnecessary because persuading the Court that the purpose behind the law is discriminatory forecloses the government's ability to show a compelling purpose for it.").

156. Students for Fair Admissions, Inc. v. President & Fellows of Harvard College, 600 U.S. 181, 206–07 (2023) (internal citations omitted).

157. *See, e.g.*, Trump v. Hawaii, 138 S. Ct. 2392, 2419 (2018) ("For another, 'when it comes to collecting evidence and drawing inferences' on questions of national security, 'the lack of competence on the part of the courts is marked.'"), *quoting* Holder v. Humanitarian Law Project, 561 U.S. 1, 34 (2010); Robert M. Chesney, *National Security Fact Deference*, 95 VA. L. REV. 1361 (2013).

On the other hand, Chief Justice Roberts recently wrote that at least with respect to any "race-based" equal protection claims, outside the circumstances of affirmative action (and the repudiated *Korematsu*), the Court has "identified only two compelling interests" that permit disparate treatment: "remediating specific, identified instances of past discrimination" and "avoiding imminent and serious risks to human safety in prisons, such as a race riot."[158] Neither seems applicable to national security or AI claims. A court convinced of disparate treatment and applying strict scrutiny could invalidate an AI-based government program.

c) It is unclear after *Trump v. Hawaii* whether the standard will be strict scrutiny or rational basis review in national security cases.

Some discriminatory government AI programs may not receive strict scrutiny. Rather, even if courts reach the merits, they might, after the 2018 travel ban case *Trump v. Hawaii*, apply rational basis review to some types of national security cases. In *Trump v. Hawaii*, plaintiffs alleged religious discrimination under the Establishment Clause of the First Amendment, and the Court applied rational basis review. (The Court first concluded that it need apply only the *Mandel* standard, which asks "only whether the policy is facially legitimate and bona fide," but determined that it could look behind the government's stated facially neutral policy for discrimination because the government had conceded as much in briefing.[159]) Justice Sotomayor, joined by Justice Ginsberg, wrote in dissent that precedent called for a more "stringent" standard of heightened scrutiny for First Amendment discrimination claims.

It is unclear how widely the Court will apply a lower standard of review: whether (1) only when the denial of visas of foreign nationals "allegedly burdens the constitutional rights of a U.S. citizen," justifying "circumscribed judicial inquiry,"[160] or (2) to all "admission and immigration cases that overlap with 'the area of national security'" where *Mandel's* narrow standard of review "has particular force";[161] or (3) (perhaps synonymous with 2) to any case where the President needs "to respond to changing world conditions," such that the Court's "inquiry into matters of entry and national security is highly constrained,"[162] or even more broadly, to national security and foreign affairs

158. *Students for Fair Admissions, Inc.*, 600 U.S. at 207.

159. *Trump*, 138 S. Ct. at 2420.

160. *Id.* at 2419.

161. *Id.*

162. *Id.* at 2418 ("[P]laintiffs seek to invalidate a national security directive regulating the entry of aliens abroad. Their claim accordingly raises a number of delicate issues regarding the scope of the constitutional right and the manner of proof. The Proclamation, moreover, is

generally. In a footnote, Chief Justice Roberts suggested he might apply "a more constrained standard of review" even to "immigration policies, diplomatic sanctions, and military actions," or even simply "the national security and foreign affairs context," though perhaps only if the entry of foreign nationals is involved.[163]

Dean Erwin Chemerinsky writes that the justices disagreed "over the extent of judicial deference to executive decisions in immigration when there is strong evidence of religious animus" and that "[t]he greatest long-term significance of the case is likely to be in the majority's using only rational basis review for scrutinizing presidential decisions in this area."[164] Hopefully, we can read *Trump v. Hawaii's* lower standard of judicial deference as applying not to all national security law cases but only to the narrower (but still important) category of immigration-related religious discrimination cases. That reading is supported by Chief Justice Roberts continuously caveating or pairing the broader "national security" with narrowing terms like "entry" and "immigration" and "foreign nationals." Further, in distinguishing *Korematsu* from *Trump v. Hawaii*, the Chief Justice wrote, "it is wholly inapt to liken" the infamous 1940s military order requiring "the forcible relocation of U.S. citizens in concentration camps," to "a facially neutral policy denying certain foreign nationals the privilege of admission. The entry suspension is an act that is well within executive authority."[165]

---

facially neutral toward religion. Plaintiffs therefore ask the Court to probe the sincerity of the stated justifications for the policy by reference to extrinsic statements—many of which were made before the President took the oath of office. These various aspects of plaintiffs' challenge inform our standard of review.").

163. *Id.* at n.5 ("The dissent finds 'perplexing' the application of rational basis review in this context. But what is far more problematic is the dissent's assumption that courts should review immigration policies, diplomatic sanctions, and military actions under the *de novo* 'reasonable observer' inquiry applicable to cases involving holiday displays and graduation ceremonies. The dissent criticizes application of a more constrained standard of review as 'throw[ing] the Establishment Clause out the window.' But as the numerous precedents cited in this section make clear, such a circumscribed inquiry applies to any constitutional claim concerning the entry of foreign nationals. The dissent can cite no authority for its proposition that the more free-ranging inquiry it proposes is appropriate in the national security and foreign affairs context.") (internal citations omitted); *see also* New York v. United States Department of Commerce, 351 F.Supp.3d 502, 666 (S.D.N.Y. 2019) ("*Trump v. Hawaii* involved review of a presidential order that 'prevent[s] the entry of [certain] *foreign nationals*' to the United States . . . It held that judicial 'inquiry into *matters of entry and national security* is highly constrained' because '[a]ny rule of constitutional law that would inhibit the flexibility of the President to respond to changing world conditions should be adopted only with the greatest caution . . . Nothing in the opinion indicates that this 'circumscribed inquiry' applies outside of the "national security and foreign affairs context.") (internal citations omitted).

164. CHEMERINSKY, *supra* note 141, at 438.

165. *Trump*, 138 S. Ct. at 2423.

But even if the lower standard of review is limited to entry and immigration cases, perhaps in a national security and First Amendment context, that is a substantial setback for civil rights. While *Trump v. Hawaii* might not supplant cases like *Hassan v. City of New York*, where the rights of U.S. citizens were affected, it might affect cases of religious discrimination at airports and at the border, where Fourth Amendment border search doctrine is already not protective. *Iqbal* and *Ziglar v. Abassi*, for example, involved foreign nationals suing on First Amendment and Equal Protection grounds. Would their allegations have received strict scrutiny or rational basis review, had they been able to successfully plead disparate treatment? Perhaps rational basis review on First Amendment claims and strict scrutiny on race-based claims? What if the government had relied on an AI program to recommend whom to detain? What if an AI recommended that all Buddhists, and only Buddhists, entering the country be flagged for searches of their luggage, laptops, and phones?

d) The impact of the recent affirmative action case on AI modeling is unclear.

The implications of the 2023 case rejecting affirmative action, *SFFA v. Harvard College*,[166] for AI are unclear. In *SFFA*, the Court embraced what scholars have referred to as an "anticlassification (formal equality)"[167] theory of equal protection law. As Professor Reva Siegel explains, scholars have often described the Justices in disagreement about whether the Equal Protection Clause should be interpreted through "a colorblind anticlassification principle concerned with individualism" or "an antisubordination principle concerned with inequalities in group status."[168] One of the most challenging aspects of AI is whether it is possible to create less biased AI, which data engineers and machine learning experts might do by explicitly adding or removing features. It remains to be seen how the Court will address this unique, technical aspect of AI. The Court's doctrinal focus on anticlassification and intent, as Professor

166. *See generally* Students for Fair Admissions, Inc. v. President & Fellows of Harvard College, 600 U.S. 181 (2023).

167. Jason R. Bent, *Is Algorithmic Affirmative Action Legal?*, 108 GEORGETOWN L.J. 803, 852 (2020).

168. Reva B. Siegel, Abstract, *From Colorblindness to Antibalkanization: An Emerging Ground of Decision in Race Equality Cases*, 120 YALE L.J. 1278 (2011). Professor Siegel goes on to posit an additional theory, the "antibalkinization principle," embraced by some justices "concerned with social cohesion, a concern analytically distinct from the value of individualism associated with colorblindness and the concern to remedy group inequality associated with antisubordination." *Id.* at 1282.

Aziz Huq observes, does not seem well suited to capturing the problems created by bias in machine learning.[169]

The issue arises when AI programmers see or foresee bias in their models and move to correct it. As Professor Huq argues, given the many ways that racism is encoded into our social norms and the resulting "lopsided diminishment in life chances and material goods for historically marginalized groups," proxies, such as zip codes and employment data, will drive machine-learning outcomes.[170] Avoiding such harms will likely require "conscientious consideration of the specific mechanisms whereby disadvantage is transmitted over time and space, and (at times) race conscious interventions to disrupt these mechanisms' operation."[171]

The Court, however, may not be sympathetic to bias-reduction work if that work is viewed more like "affirmative action" for AI.[172] Rather, it might require AI models to be run on "neutral" data. That neutrality would be a myth, especially if historic and current systematic injustices were reflected in the training, testing, and validation data.

In equal protection challenges to discriminatory AI, this issue might show up as a defense.[173] The government might argue that had it removed the bias by "affirmative action" AI, it would have violated the anticlassification theory of equal protection law embraced by *SSFA*. Therefore, it went ahead with the biased AI. One response to that defense is that knowing the AI was biased, the government *could have refrained from using the AI at all*, or it could have started anew with different training data and algorithms. In choosing to use biased AI, knowing it would produce discriminatory results, the government effectively (and intentionally) discriminated.

But perhaps *SFFA* can be read and applied the opposite way: against the original, biased AI.[174] If the biased AI bins or targets people by suspect category—that is, if it is more likely to categorize a person of color as a "risk" than it is to categorize a white person as a risk, then it is making facial classifications subject to strict scrutiny. The original, biased AI is the facially

169. Aziz Z. Huq, *Constitutional Rights in the Machine-Learning State*, 105 CORNELL L. REV. 1875, 1923–27 (2020).

170. *Id.* at 1926.

171. *Id.*

172. Associate Dean Bent, *supra* note 167, wrestles with this problem in the employment law context, ultimately finding ways to justify progressively "race-aware" algorithms.

173. It might also show up as an anticlassification claim by individuals from a non-marginalized group that race-conscious modeling to avoid biased outputs disadvantaged them.

174. I tend to agree with Professor Huq that current doctrinal equal protection law is ill suited to AI. But within the constraints of that current doctrine, I offer this potential reading of SFFA's implications for AI.

discriminatory policy. Then, any data engineering or modeling done to reduce bias would be corrective, like desegregation.[175] Reducing bias would look less like affirmative action, which the Court has rejected, and more like an anticlassification tool.

Essentially, this argument asks courts to look at both the algorithm and human conduct at a specific moment in time: the moment of the government's decision to use the final AI model, rather than the initial design and programming phases. In examining the AI, rather than looking only at the inputs and design, which may or may not show bias, it asks the court to look at the outputs, that is, the final product or version of the AI.[176] Rather than looking at the initial "neutral" programming, which is never, in fact neutral, courts should look at the moment the government determines to whether to proceed with using the biased AI. At that point, if the government were to proceed, it would create a facially discriminatory policy or government program. Any steps the government then takes to mitigate bias are intended to achieve a more facially neutral program.

e) Injunctive relief may not be available in the moment.

A final and major problem with injunctive relief is that it may not be available at the time it is needed. The plaintiffs in *Ziglar v. Abbasi*, for example, who sued for monetary damages after-the-fact of their alleged violent abuse during confinement, did not appear to have been able to sue at the time it was happening. The plaintiffs were among eighty-four "aliens" held in the Metropolitan Detention Center (MDC) in Brooklyn while the FBI investigated them post 9/11; they alleged, among other things, that the government held them in harsh conditions of confinement "because of their actual or apparent race, religion, or national origin, in violation of the equal protection component of the Fifth Amendment" and without any evidence of their involvement in terrorism.[177] Plaintiffs alleged that MDC "guards slammed [them] into walls; twisted their arms, wrists, and fingers; broke their bones; referred to them as terrorists; threatened them with violence; subjected them to humiliating sexual comments; and insulted their religion.[178]

The Court accepted those allegations as true for purposes of its decision to deny *Bivens* relief after-the-fact, even as it recited allegations that there was little or no opportunity to lawyer up and sue for injunctive relief at the time of

175. *See Students for Fair Admissions, Inc.*, 600 U.S. at 207.

176. With the caveat that this "final version" is always changing, and the government must therefore continually evaluate whether it remains constitutional to use the AI.

177. Ziglar v. Abbasi, 582 U.S. 120, 129 (2017).

178. *Id.* at 128.

the abuse. Justice Kennedy summarized the allegations of "harsh" conditions in the detention unit: detainees were "held in tiny cells for over 23 hours a day" with the lights on 24 hours a day, "[t]hey were denied access to most forms of communication with the outside world," and they were strip searched often.[179] Justice Breyer wrote in dissent that some plaintiffs alleged that they were subjected to a "communications blackout" for two to three months; that prison staff denied them visitors, legal and social telephone phone calls, and mail; that neither their families nor attorneys knew where they were being held; "that they could not receive visits from their attorneys; that subsequently their lawyers could call them only once a week; and that . . . defendants interfered with the detainees' effective access to legal counsel."[180] Justice Breyer emphasized that neither prospective injunctive relief nor a writ of habeas corpus "will normally provide plaintiffs with redress for harms they have already suffered."[181]

While AI will not hold anyone in prison nor beat them (at least, not in near future), it might well be used to help determine whom to investigate, arrest, and detain. Likewise, AI might be used to determine whose neighborhoods and places of worship or affiliation to surveille, and whose speech to monitor on social media. But as in *Ziglar*, injunctive relief may not be available in the moment, nor will it address past harms. After-the-fact monetary damages must be available.

C. *BIVENS* PRESENTS NO VIABLE OPTION FOR POST-HARM RELIEF FOR DISCRIMINATORY AI.

Without a statutory cause of action, plaintiffs will not be able to sue for after-the fact damages for discriminatory AI in national security cases. The judicially-created remedy of *Bivens* damages for constitutional violations by federal officers is a non-starter in national security contexts (and perhaps in any context). In *Bivens v. Six Unknown Fed. Narcotics Agents*, in the absence of a statutory cause of action, the Supreme Court held that the Fourth Amendment provided "a damages remedy against for those whom federal officers have injured as a result of an unconstitutional search or seizure."[182] Two subsequent cases found an implied damages remedy under the Fifth Amendment in an equal protection claim and under the Eighth Amendment's prohibition against cruel and unusual punishment.[183] But

179. *Id.*
180. *Id.* at 173 (internal quotations omitted).
181. *Id.*
182. *Id.* at 129 (2017) (Breyer, J., dissenting) (summarizing *Bivens*).
183. *Id.*

recent cases, including *Ziglar v. Abbasi*, have marked the "decline, if not death, of *Bivens*."[184]

Despite the horrendous nature of the allegations in *Ziglar v. Abbassi*, described above, the Supreme Court effectively foreclosed any prospect for post-harm *Bivens* relief in national security cases: "a 4-2 majority of the Supreme Court expressed serious skepticism that *Bivens* claims for damages will ever be appropriate in the context of national security."[185] Justice Kennedy wrote that separation of powers concerns cautioned against imposing after-the-fact, monetary liability on national security officials because such liability might cause them to "second-guess difficult but necessary decisions concerning national-security policy."[186] The 2022 Supreme Court decision *Egbert v. Boule* effectively foreclosed national security *Bivens* cases.[187] Justice Thomas opined, too, that the analysis of whether to provide a judicial *Bivens* remedy "often resolve[s] to a single question: whether there is any reason to think that Congress might be better equipped to create a damages remedy."[188]

A significant problem with having no monetary damages for constitutional violations is that it takes away a key incentive for government officials to abide by the law.[189] Government actors worry about financial liability and going to jail. While officials do have other incentives for following the Constitution, such as maintaining individual self-respect and morality or maintaining the government or agency's mission and reputation, imposing personal liability for government abuses would be a more effective method of achieving accountability. Holding an agency or group responsible via an injunction provides some incentive to uphold the law—most employees likely wouldn't want to create a program only to have it ordered shut down—but singling out individuals for liability is a much bigger stick.

With respect to AI, accountability is of course challenging: who should be held liable for discriminatory AI, especially when the AI may become discriminatory over time as it encounters new data? Is it the designer, the programmer, the official who requests it, the official who approves it, the contractor who provides or maintains it, the employee who uses and interprets it, or some combination of those individuals? I would argue that at a minimum, any contractor who provides and profits from a discriminatory model should

184. HOWARD M. WASSERMAN, UNDERSTANDING CIVIL RIGHTS LITIGATION 106 (3d ed. 2023).

185. STEPHEN DYCUS, WILLIAM BANKS, EMILY BERMAN, PETER RAVEN-HANSEN & STEPHEN I. VLADECK, SUPPLEMENT TO NATIONAL SECURITY LAW 150 (7th ed. 2022).

186. *Ziglar*, 582 U.S. at 142; *see* Dycus et al., *supra* note 185, at 150–51.

187. Egbert v. Boule, 596 U.S. 482 (2022); *see* Dycus et al., *supra* note 185, at 150–51.

188. *Egbert*, 596 U.S. at 492.

189. Dycus et al., *supra* note 184, at 150–51.

be accountable. Like the government, the contractor has an obligation to test and monitor its AI at least up until the point of sale (and possibly afterward, if the contractor administers the government program or finds bias in its models employed in other contexts). Government officials who request or approve of the use of discriminatory AI should be liable.[190] Decisions to use AI should be made at the highest levels because of their grave consequences for humanity.

In light of the Court's stated deference to the political branches,[191] Congress should consider creating statutory causes of action for damages for AI discriminatory harms, as discussed later in Section V.A.

D. WATCHLISTING: AI RISKS FOR FIFTH AMENDMENT DUE PROCESS

Government use of biased AI in national security watchlisting, surveillance, and any criminal, counterintelligence, or immigration investigations raises significant Fifth Amendment due process issues. Watchlisting cases provide an illustrative example. While the government might have strong incentive to use AI for watchlisting—it might help sort needles from the haystack—doing so poses significant risk of profiling and error. As noted in Part III, watchlists are already populated predominately by Muslims; it is difficult to imagine an unbiased historical data training data set. (Similarly, current biases, such as, perhaps, any informed by great power competition with China and Russia, might inform surveillance or investigatory decisions.) Government watchlists are also notorious for their error rates.[192] AI will exacerbate these issues.

In cases not addressing AI, some courts have found due process violations in the nomination process for government watchlists and in the government's redress process for individuals unjustly denied or delayed flight boarding.[193] Applying the *Mathews v. Eldridge*[194] three-factor test, courts have considered the individual's liberty interest, including the right to travel and to be free from

190. Problematically, *Iqbal* rejected the idea of supervisory liability in the Bivens context, all the more reason for a Congressional remedy.

191. *Ziglar*, 582 U.S. at 135–36; *Egbert*, 596 U.S. at 502.

192. *See* Letter to Executive Officials from 13 Senators and Members of Congress, *supra* note 48, at 3–4.

193. *E.g.*, Ibrahim v. Department 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/.

194. Mathews v. Elridge, 424 U.S. 319, 335 (1976).

incarceration and the stigma of being denied boarding or watchlisted;[195] the risk of erroneous error and probable value of additional or substitute procedural safeguards; and the government's security interest in watchlisting. Where courts have determined that an individual's liberty interest has been infringed, cases have turned on the second factor, the risk of erroneous error and the probable value of additional or substitute procedural safeguards.[196]

In *Ibrahim*, for example, an FBI agent mistakenly nominated the plaintiff to the No-Fly list by marking a checklist form in exactly the opposite way it was intended.[197] The Northern District of California held that due process required a correction of "the error and all of its echoes" in all government records and "interlocking databases."[198] If an AI system were to make a similar mistake, that error might not be readily transparent, if it were a decision made within the black box, or based on multiple inputs. Nor might it be easily corrected. In a world of interwoven data sets and government AI systems, ridding all echoes of the mistake might be a very challenging remedy for the government to implement. Further, as noted in a recent Senate committee report, the extent to which terrorist watchlist information is shared outside the federal government, with state, local, and tribal governments, and select international partners and private entities, is unclear.[199] Due process, however, may require such corrections.

In *Latif v. Holder*, 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.[200] The notice had to be reasonably calculated to permit plaintiffs to submit evidence rebutting the government's reasons for their

195. *Compare Ibrahim*, 62 F. Supp. 3d at 928 *and Latif*, 28 F. Supp. 3d at 1148–51 *with Elhady*, 993 F.3d at 226–07, *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).

196. *See Latif*, 28 F. Supp. 3d at 1160–61; *Ibrahim*, 62 F. Supp. 3d at 929; *see also* STEPHEN DYCUS, WILLIAM BANKS, EMILY BERMAN, PETER RAVEN-HANSEN & STEPHEN I. VLADECK, NATIONAL SECURITY LAW (7th ed. 2019) Teachers' Manual, 25-7; *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).

197. *Ibrahim*, 62 F. Supp. 3d at 928.

198. *Id.* at 929.

199. S. COMM. ON HOMELAND SECURITY & GOVERNMENTAL AFFAIRS, MAJORITY STAFF REPORT: MISLABELED AS A THREAT: HOW THE TERRORIST WATCHLIST & GOVERNMENT SCREENING PRACTICES IMPACT AMERICANS 31–32 (Dec. 2023), https://www.hsgac.senate.gov/wp-content/uploads/Mislabeled-as-a-Threat_Public_Report-2.pdf.

200. *Latif*, 28 F. Supp. 3d at 1162.

inclusion on the list.[201] With AI in the mix, a court might have to require the government to make its algorithm available in discovery, or at least to reveal those inputs and factors the algorithm considered in nominating a particular individual to the list.

Both *Latif* and the district court in *Elhady*[202] noted the low standard—the executive's reasonable suspicion standard—for inclusion on the lists. In *Elhady*, which was later reversed, the Eastern District of Virginia determined that the central national database from which all other, shorter lists are derived, the TSDB, posed due process issues. The court cited the vague and low standard for including an individual on the TSDB, noting the plaintiffs' assertions that the Terrorist Screening Center "may consider a wide range of factors in determining whether an individual belongs on the Watchlist, including an individual's 'race, ethnicity, or religious affiliation,' beliefs and activities protected by the First Amendment, travel history, personal and professional associations, and financial transactions."[203] Moreover, the court found "there is no independent review of a person's placement on the TSDB by a neutral decisionmaker," which "coupled with the limited disclosures and opportunity to respond by a person who requests that his status be reviewed," creates a substantial risk of erroneous deprivation.[204] Lest one becomes too optimistic about the prospect of judicial review of watchlisting cases, I note that the Fourth Circuit reversed the district court's finding of a due process violation in *Elhady*, on the ground that the plaintiffs' travels were delayed but not precluded.[205]

The low standard used in the watchlists nomination process remains in place, and it will be subject to greater error and bias in the age of AI. Any courts willing to review watchlist determinations will need to be ready to review classified and possibly proprietary systems and algorithms, or at a minimum specific inputs and outputs. It is easy to imagine some judges willing to insist on such review, and others deferring to government claims of secrecy, national security imperatives, and a lack of technical expertise.

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

202. Elhady v. Kable, 391 F. Supp. 3d 562 (E.D. Va. 2019).

203. *Id.* at 581 (citing Pls.' Statement of Material Facts).

204. *Id.* at 581–82.

205. Elhady v. Kable, 993 F.3d 208 (4th Cir. 2021).

E. OTHER BARRIERS TO RELIEF: STANDING, STATE SECRETS, QUALIFIED IMMUNITY

Traditional barriers to national security challenges—justiciability doctrines, the state secrets doctrine and classification more broadly, qualified immunity[206]—may be heightened in cases of potentially discriminatory AI programs. The AI itself will almost necessarily be classified; some of it may be proprietary to government contractors as well. Secrecy will make it especially challenging for plaintiffs to show injury for standing. That was true, for example, in the non-AI cases *Clapper*,[207] where plaintiffs could not establish standing because the Court thought their injuries under the classified 702 surveillance program too speculative, and *Halkin v. Helms*,[208] where the government asserted states secrets, blocking discovery necessary for plaintiffs to assert standing for their claims about unlawful surveillance during the Vietnam War.

Courts might review both classified and proprietary materials in camera. Due process almost certainly requires that eligible counsel be given clearance to do the same in many or all cases. Courts might insist about disclosure of both classified and proprietary materials if government programs are to go forward. Congress might likewise legislate that any AI applications that are likely to impact civil or human rights likewise be publicly, rather than privately, owned, even if they are to be classified. But whether courts will insist on such due process protections is questionable given the history of national security cases being dismissed on justiciability and secrecy grounds.

The qualified immunity doctrine, too, may block constitutional challenges: before government officials are held liable for constitutional violations, plaintiffs must show that the officials violated a "clearly established" constitutional or statutory right. While equal protection, due process, and First and Fourth Amendment rights are all clearly established, it is easy to imagine a court ducking enforcing those rights by suggesting that AI's impact on those rights is largely untreated by case law and statute, and thus the law is not clearly established. This is yet another reason for Congress to legislate and help resolve AI questions. As recommended below, Congress should provide causes of action for damages relief for disparate treatment and injunctive and damages relief for disparate impact. It might provide statutory causes of action for damages for First, Fourth, and Fifth Amendment violations.

206. *See generally* Dycus et al., *supra* note 60, ch. 5.
207. *See generally* Clapper v. Amnesty Int'l, 568 U.S. 398 (2013).
208. Halkin v. Helms, 690 F.2d 977 (C.A.D.C. 1982).

IV. EXECUTIVE POLICIES RELEVANT TO DISCRIMINATORY AI ARE TOO PERMISSIVE

This section reviews primary executive branch policies relevant to AI and national security surveillance and investigations. While they have some good aspects to them—in particular, some of the more recent AI-specific policies show a technical understanding of algorithmic discrimination—they also allow generous room for secrecy and discretion by government officials, and they are all reversible by the new administration. Without increased, interbranch oversight, any executive policy will have a significant weakness: it will be executed behind classified doors. Accountability will only be as good as the ethics and fortitude of the individuals in the room. And unconscious biases, including limited experiential perspective if the room insufficiently diverse, will be even harder to unmask, even for good faith actors.

A. DEPARTMENT OF JUSTICE GUIDANCE FOR FEDERAL LAW ENFORCEMENT AGENCIES ON THE USE OF SUSPECT CLASSIFICATIONS

In May 2023 the Department of Justice updated its "Guidance Regarding the Use of Race, Ethnicity, Gender, National Origin, Religion, Sexual Orientation, Gender Identify, and Disability."[209] As Professor Faizel Patel writes, while the 2023 Guidance includes some improvements, such explicitly addressing intelligence activities supporting law enforcement, it "continue[s] to allow targeting based on group characteristics rather than indications of individual wrongdoing—the very essence of the invidious profiling the rules claim to ban."[210]

AI complicates this by obscuring the discriminatory bias within the workings of the machine and by magnifying its spread—as discussed above, by increasing the net of surveillance and potentially linking data in multiple types of databases (biometric data with tax and loan data, for example) and, as Professor Hu has argued, between law enforcement and national security data regimes.[211]

209. U.S. DEP'T OF JUSTICE, *Guidance*, *supra* note 8.

210. Faiza Patel, *Threat from Within? Unreformed Counterterrorism Infrastructure Raises Concerns About Misuse*, JUST SECURITY (Nov. 21, 2023), https://www.justsecurity.org/90142/threat-from-within-unreformed-counterterrorism-infrastructure-raises-concerns-about-misuse/; *see also* Faiza Patel & Hina Shamsi, *DOJ and DHS Racial Profiling Guidelines Must Close Loopholes Permitting Bias*, JUST SECURITY (May 15, 2023), https://www.justsecurity.org/86577/doj-and-dhs-racial-profiling-guidelines-must-close-loopholes-permitting-bias/ (anticipating the release of the 2023 Guidelines).

211. Hu, *supra* note 61.

To play out an example of how AI can exacerbate harms, here is a scenario from the Guidance:

> To undertake a national or homeland security operation, or an intelligence action based on a listed characteristic, law enforcement personnel must have trustworthy information that contains context- and content-specific details linking persons possessing that characteristic to a threat to national or homeland security, or intelligence authorized activity, and the actions undertaken must be reasonable under the totality of circumstances.
>
> • Example: A Federal law enforcement agency receives reliable information that persons affiliated with a foreign ethnic insurgent group intend to use hand-delivered explosive devices to assassinate that country's president and his entire entourage during an official visit to the United States. Agents may appropriately focus investigative attention on identifying members of that ethnic insurgent group who may be present and active in the United States and who, based on other available information, might be involved in planning some such attack during the state visit.[212]

In this example, it appears that federal agents might seek to identify members of the ethnic insurgent group present in the United States, based on their ethnicity, and possibly "other available information" suggesting they might be involved in planning the attack. Ethnicity and perhaps national origin are the key factor(s) in focusing "investigative attention." *Perhaps* some behavioral factor might be included (the "other available information") linking an individual to the attack plan. However, other than the suggestion that the attackers will use hand-delivered explosives, the tip does not provide much to go on to establish any individually-behavior based grounds for identifying possible suspects.

If the government investigators were to employ AI to search for people who might be involved in the "foreign ethnic insurgent group" using ethnicity and national origin, the results would look like the propensity profiling that the *Farag* court considered unconstitutional.

It seems the government is betting on these scenarios not being litigated on the merits, and if they are, that they will be able to survive strict scrutiny by deferential courts.

212. U.S. DEP'T OF JUSTICE, *Guidance*, *supra* note 8, at 12.

B. INTELLIGENCE COMMUNITY PRINCIPLES AND ETHICAL FRAMEWORK FOR AI

The 2020 Artificial Intelligence Ethics Framework for the Intelligence Community[213] offers excellent and detailed questions for technologists and attorneys to ask before and while using any AI tool. However, in allowing room for flexibility, it also allows room for civil rights and civil liberties to be outweighed by security interests. This balancing, too, is to be done by government officials in classified environments, rather than by courts or Congress in the public eye. (Even intelligence community regulations are classified and do not go through the Federal Registrar notice and comment and publication.)

For example, under the heading of "Purpose: Understanding Goals and Risks," the document suggests that executive officials "determine what goals you are trying to achieve to ensure you can design AI that balances desired results with acceptable risk," by asking a series of questions:

> What is the goal you are trying to achieve by creating this AI . . . ? Is there a need to use AI to achieve this goal? Can you use other non-AI related methods to achieve this goal with lower risk? Is AI likely to be effective in achieving this goal?
>
> Are there specific AI system methods suitable and preferred for this use case? Does the efficiency and reliability of the AI in this particular use case justify its use for this purpose?
>
> What benefits and risks, including risks to civil liberties and privacy, might exist when this AI is in use? Who will benefit? Who or what will be at risk? What is the scale of each and likelihood of the risks? How can those risks be minimized and the remaining risks adequately mitigated? Do the likely negative impacts outweigh likely positive impacts?[214]

It is excellent to ask questions such as whether the AI is necessary at all, and whether the specific AI is well suited to the task at hand.[215] Likewise, to

213. OFF. OF THE DIR. OF NAT'L INTEL., ARTIFICIAL INTELLIGENCE ETHICS FRAMEWORK FOR THE INTELLIGENCE COMMUNITY (Jan. 15, 2023), https://www.intelligence.gov/artificial-intelligence-ethics-framework-for-the-intelligence-community.

214. *Id.*

215. In the criminal justice system, for example, algorithms designed to predict parole risk for future crime are not at all suited to determining punishment for past crimes, but courts have nonetheless employed them at sentencing (*see* Loomis v. Wisconsin, 881 N.W.2d 749 (Wis. 2016), *cert. denied*, 137 S. Ct. 2290 (2017)); there are of course many other issues with algorithms used for any purpose in the criminal justice system, a key one being that they are inevitably discriminatory; my own view is that they should not be used at all.

ask who will benefit and who or what will be at risk. (It would be preferable yet to consult stakeholder groups.) But it is too much to expect that executive officials and attorneys, who have "mission"-driven values and imperatives along with their own socially produced biases and agency-produced biases, will be able to "balance" the risks and determine whether and how risky AI should be used. It is unrealistic to expect that they will perfectly understand and account for the potential harm to particular groups. It is also unrealistic to expect an agency attorney, who already has an endlessly busy job of dealing with day-to-day agency operations, administration, litigation, etc., to be the gatekeeper of the use and maintenance of complex AI systems. But right now, those attorneys, the officials they advise, and perhaps a few civil rights and civil liberties and Inspector General offices are about all we've got. The Biden Administration's addition of Chief AI Officers and Governance Boards, discussed below, was certainly a step in the right direction; hopefully those positions will continue to exist. Congressional action is still required.

C. EXECUTIVE ORDERS ON AI AND IMPLEMENTING MEMORANDA:

In October 2023, the Biden Administration released Executive Order (EO) 14110 on Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence.[216] In late January 2025, just before this Article was to be published, the Trump Administration revoked[217] EO 14110, replacing it with a much briefer Executive Order 14179, "Removing Barriers to American Leadership in Artificial Intelligence."[218] The Trump Order directs that any policies and directives taken pursuant to the Biden Order will be reviewed "immediately" for inconsistency with "the policy of the United States to sustain and enhance America's global dominance in order to promote human flourishing, economic competitiveness, and national security."[219]

Because the Biden Administration's efforts represent the most significant to date at the federal level toward regulating AI, including addressing AI bias, and because the implementing policy memorandum directed by the Biden Order appear to remain at least temporarily in place, I analyze those policies here. While the Trump Administration is likely to significantly revise or revoke the Biden policies, they still provide a useful starting point for future government action. They are imperfect, as I will suggest, but certainly more

216. Exec. Order No. 14,110, 88 Fed. Reg. 75,191 (2023).

217. Exec. Order 14,148, 90 Fed. Reg. 8,237 (2025) (rescinding Biden EO 14110 on AI, among other Biden Executive Orders).

218. Exec. Order No. 14,179, 90 Fed. Reg. 8,741 (2025).

219. Exec. Order No. 14,179 §§ 5, 2, 90 Fed. Reg. at 8,741.

protective than a vacuum of regulation. And their imperfections can be remedied by a future Order, or better yet, a statute.

The Biden Executive Order and several implementing memoranda and policy documents demonstrate a deep understanding of AI technologies and risks. They represent a real step forward in regulating AI use by government actors. They also, however, establish a potentially problematic bifurcation between how national security AI uses and non-national security AI uses are to be governed. As Professor Patel and the ACLU's Patrick Toomey have suggested, having two separate regulatory systems may lead to less protection of civil and human rights in national security contexts.[220]

The bifurcation began in the now-revoked Executive Order. Although EO 14110 established "initial means, instructions, and guidance" for addressing government use of AI, those sections did "not apply to AI . . . used as a component of a national security system."[221]

With respect to *non*-national security systems, EO 14110 directed the Office of Management and Budget (OMB) to issue guidance for government use of AI. The guidance was to address best practices such as "conducting public consultation; assessing data quality; assessing and mitigating disparate impacts and algorithmic discrimination; providing notice of the use of AI; continuously monitoring and evaluating deployed AI; and granting human consideration and remedies for adverse decisions made using AI."[222] After a public comment period, OMB released a final policy Memorandum, M-24-10, on March 28, 2024.[223]

With respect to AI used on national security systems and for military or intelligence purposes, EO 14110 called for an interagency process to develop a "National Security Memorandum on AI."[224] The White House published an unclassified National Security Memorandum on AI ("AI NSM") on October 24, 2024, along with a classified annex.[225] The AI NSM was implemented in

220. Faiza Patel & Patrick C. Toomey, *National Security Carve-Outs Undermine AI Regulations*, JUST SECURITY (Dec. 21, 2023), https://www.justsecurity.org/90771/national-security-carve-outs-undermine-ai-regulations/.

221. Exec. Order No. 14,110 § 10.1(i), 88 Fed. Reg. at 75,191.

222. Exec. Order No. 14,110 § 10.1(b)(iv), 88 Fed. Reg. at 75218.

223. OMB MEMORANDUM M-24-10, ADVANCING GOVERNANCE, INNOVATION, AND RISK MANAGEMENT FOR AGENCY USE OF ARTIFICIAL INTELLIGENCE (Mar. 28, 2024), https://bidenwhitehouse.archives.gov/wp-content/uploads/2024/03/M-24-10-Advancing-Governance-Innovation-and-Risk-Management-for-Agency-Use-of-Artificial-Intelligence.pdf.

224. Exec. Order No. 14,110 § 4.8, 88 Fed. Reg. at 75204.

225. THE WHITE HOUSE, AI MEMORANDUM ON ADVANCING THE UNITED STATES' LEADERSHIP IN ARTIFICIAL INTELLIGENCE; HARNESSING ARTIFICIAL INTELLIGENCE TO FULFILL NATIONAL SECURITY OBJECTIVES; AND FOSTERING THE SAFETY, SECURITY, AND

part by another document, a "Framework to Advance AI Governance and Risk Management in National Security" ("NS AI Framework").[226]

As of early February 2025, the OMB Memorandum M-24-10, the AI NSM, and the NS AI Framework appear to remain in effect, subject to review and possible revision or revocation by the Trump Administration.[227]

From a civil rights perspective, comparing the NS AI Framework to the OMB Memorandum M-24-10 is a bit like comparing a car with base-level trim to one with mid-level trim: while many of the features are similar, the base level lacks several significant features. The fact that the government decided to create two distinct but overlapping frameworks suggests that any differences between them are intentional and important. Here, I will examine salient features and differences with an eye toward improving future governance of national security AI.

Internal Agency Governance. Under both frameworks, agencies are required to designate Chief AI Officers and AI Governance Boards to advise on AI and institute governance and oversight.[228] This is a helpful initiative that should be preserved.

Covered AI Uses. Both frameworks also require agencies to institute "minimum risk management practices" before using certain categories of AI applications. For AI covered by the OMB Memorandum M-24-10 framework,

---

TRUSTWORTHINESS OF ARTIFICIAL INTELLIGENCE ("AI NSM") (Oct. 24, 2024), https://bidenwhitehouse.archives.gov/briefing-room/presidential-actions/2024/10/24/memorandum-on-advancing-the-united-states-leadership-in-artificial-intelligence-harnessing-artificial-intelligence-to-fulfill-national-security-objectives-and-fostering-the-safety-security/.

226. THE WHITE HOUSE, FRAMEWORK TO ADVANCE AI GOVERNANCE AND RISK MANAGEMENT IN NATIONAL SECURITY ("NS AI Framework") (Oct. 24, 2024), https://ai.gov/wp-content/uploads/2024/10/NSM-Framework-to-Advance-AI-Governance-and-Risk-Management-in-National-Security.pdf.

227. The Trump Executive Order addressed to AI, EO 14,179, *supra* note 218, directs that OMB Memorandum M-24-10 shall be "revise[d]" "as necessary" within sixty days to be consistent with the policy goals of the Order. Exec. Order No. 14,179 § 5, 2, 90 Fed. Reg. at 8,741. Neither the AI NSM nor the NS AI Framework are specifically addressed by EO 14,179, other than in its general directive to revise and/or revoke any policies or actions made pursuant to the Biden Order on AI. *Id.* The Trump EO 14,179 on AI does direct the Assistant to the President for National Security Affairs (APNSA) (the "National Security Advisor"), along with other officials, to develop and submit to the President with 180 days an "action plan" to achieve the policy goals of the Order. *Id.* at § 4. The separate, earlier Trump Executive Order 14148, *supra* note 217, which rescinded the Biden Executive Order 14110 on AI and a host of other Biden executive orders, directed the National Security Advisor to review within 45 days all National Security Memoranda for potential rescission. Exec. Order No. 14148 § 3(c), 90 Fed. Reg. at 8241.

228. OMB Memorandum M-24-10, *supra* note 223, § 3(a); NS AI Framework, s*upra* note 226, at 9-2.

the minimum practices appear to address all "safety-impacting" or "rights-impacting" AI.[229] For AI covered by the NS AI Framework—i.e., AI to be used in National Security Systems—the minimum practices apply only to "high-impact" AI.[230] High-impact AI is defined as "AI whose output serves as a principal basis for a decision or action that could exacerbate or create *significant* risks to national security, international norms, democratic values, human rights, civil rights, civil liberties, privacy, or safety[.]"[231] Further, "AI use is presumed to be high impact if it *controls or significantly influences* the outcomes" of a non-exhaustive list of activities, including tracking or identifying individuals in real time based solely on biometric data; classifying an individual as a known or suspected terrorist or national security threat; determining an individual's immigration or asylum status; and performing criminal risk assessments and predictions inside the United States or with respect to U.S. persons, or in relation to immigration or entry into the United States.[232] This setup begs the question of whether there are lesser impact national security uses, such as perhaps where an AI output "influences" rather than "significantly influences" a government decision or action that could affect civil rights, where the minimum risk management practices do not apply.

Risk and Impact Assessments. As part of the minimum risk management practices, both the OMB Memorandum M-24-10 and the NS AI Framework require agencies to perform risk and impact assessments for new and existing AI applications before they are used.[233] Agencies must identify the intended purpose, expected benefits, and potential risks of the application.[234] Both frameworks also require agencies to test AI applications for performance in a real world context, and to seek independent testing (within the agency) of the AI before its deployment.[235]

But while the NS AI Framework closely tracks some of the OMB Memorandum M-24-10's language about risk assessments, it drops other language, such as that agencies should "document the stakeholders who will be most impacted by the use of the system."[236] Likewise, only the OMB

229. OMB Memorandum M-24-10, *supra* note 223, § 5(c)(iv).

230. NS AI Framework, *supra* note 226, at 5. Some AI uses are explicitly prohibited, including using AI "with intent or purpose" of "unlawful" discrimination.

231. NS AI Framework, *supra* note 226, at 3 (emphasis added).

232. *Id.* at 3–4 (emphasis added).

233. OMB Memorandum M-24-10, *supra* note 223, § 5(c)(iv); NS AI Framework, *supra* note 226, at 5–6.

234. *Id.*

235. OMB Memorandum M-24-10, *supra* note 223, § 5(c)(iv); NS AI Framework, *supra* note 226, at 6 (requiring independent review is only required to the "extent practicable.").

236. OMB Memorandum M-24-10, *supra* note 223, § 5(c)(iv).

Memorandum M-24-10 requires agencies to document in their risk assessments the "representativeness" of the AI's data for its intended purposes and to "assess whether the data used can produce or amplify inequitable outcomes as a result of poor data representativeness or harmful bias," including from historical discrimination.[237] The NS AI Framework later suggests that, at a systemic level, department heads shall continuously evaluate data management policies, including policies for evaluating AI training data for robustness, representativeness, and "reasonably foreseeable" bias.[238]However, it might be questioned why national security agencies should not specifically address data representativeness at the individual AI application level when performing risk assessments.

Identifying and Mitigating Algorithmic Bias. For each AI application, both the OMB Memorandum and the National Security AI Framework require that agencies seek to identify and mitigate algorithmic discrimination and disparities across groups.[239] Notably omitted from the National Security AI Framework, however, is a whole section included in the earlier OMB Memorandum M-24-10 entitled "Additional Minimum Practices for Rights-Impacting AI."[240] This section specifically requires that agencies address fairness, equity, and algorithmic discrimination in several contexts, including performing risk assessments; consulting with affected communities; monitoring AI systems post-deployment; notifying individuals negatively affected by AI outputs; and providing timely human consideration and remedies for individuals harmed by AI.[241] These protections appear to be absent from the NS AI Framework.[242]

OMB's "Additional Minimum Practices" section further requires that agencies "[i]dentify and document in their AI impact assessment when [they are] using data that contains information about a class protected by Federal nondiscrimination laws (e.g., race, age, etc.),"[243] and they "should also assess and document whether the AI model could foreseeably use other attributes as proxies for a protected characteristic and whether such use would significantly influence model performance."[244] This difference between the two frameworks is problematic. Classification and secrecy cannot explain why

237. OMB Memorandum M-24-10, *supra* note 223, § 5(c)(iv), n.37.

238. NS AI Framework, *supra* note 226, at 9.

239. OMB Memorandum M-24-10, *supra* note 223, § 5(c)(v)(A); NS AI Framework, *supra* note 226, at 6.

240. OMB Memorandum M-24-10, *supra* note 223, § 5(c)(v).

241. *Id.*

242. The NS AI Framework discusses notice and remedies only for federal personnel negatively impacted by AI. NS AI Framework, s*upra* note 226, at 7.

243. *Id.*

244. *Id.*

national security agencies using AI should not specifically consider bias, protected categories, and proxies in their risk assessments. The well-understood concept of proxies does not even appear by name in the NS AI Framework.

The OMB Memorandum requires agencies cease use of an AI application if the agency is unable to adequately mitigate the risk of unlawful discrimination.[245] The NS AI Framework only provides, more generally and without direct reference to bias or discrimination, that if the potential benefits of an AI application do not meaningfully outweigh the unmitigated risks, agencies should not use the AI.[246]

Consulting Stakeholders. OMB's "Additional Minimum Practices" also include a page of guidance about how agencies must consult and incorporate feedback from affected communities and, where appropriate, the public.[247] The NS AI Framework omits this guidance, at least as a requirement for each and every AI rights-impacting use case, presumably due to concerns about classification and secrecy.[248] The omission, however, places even greater importance on the need for transparency and for consultation with and consideration of stakeholders and affected communities wherever possible.[249]

Monitoring AI in the Field. Both frameworks discuss monitoring AI after its deployment.[250] Critically, the OMB framework provides that as part of monitoring requirements, "agencies must also monitor rights-impacting AI to specifically assess and mitigate AI-enabled discrimination against protected classes" that might arise during deployment.[251] Where "sufficient" mitigation

245. OMB Memorandum M-24-10, *supra* note 223, § 5(c)(v)(A).

246. NS AI Framework, *supra* note 226, at 5–6.

247. OMB Memorandum M-24-10, *supra* note 223, § 5(c)(v)(B).

248. The NS AI Framework does include, as part of its oversight guidance, that officials with oversight responsibility for privacy, civil liberties, transparency, and safety, might "seek and consider feedback from relevant stakeholders, including civil society, technologists, academics, the private sector, and impacted communities, as appropriate." NS AI Framework, *supra* note 226, at 11. This consultation is not required for every rights-impacting AI use case as it is in the OMB Memorandum.

249. That need is weakly contemplated by the NS AI Framework's note that to "address potential risk management gaps," covered agencies are "encouraged" to incorporate best practices from the OMB Memorandum M-24-10, the Artificial Intelligence Ethics Framework for the Intelligence Community, and other executive branch AI risk frameworks. NS AI Framework, *supra* note 226, at 5 (additionally citing DOD's Responsibility AI Strategy and Implementation Pathway, the Blueprint for an AI Bill of Rights, the NIST AI Risk Management Framework, the DHS Safety and Security Guidelines for AI in Critical Infrastructure, and EO 14,110).

250. Exec. Order No. 14,110 § 4.8, 88 Fed. Reg. at 75204.

251. OMB Memorandum M-24-10, *supra* note 223, § 5(c)(v)(C).

is not possible, agencies are required to discontinue use of the AI.[252] As the OMB Memorandum recognizes, continued monitoring for bias is critical if AI encounters different data in the real world and/or learns poor behavior. While the NS AI Framework provides that agencies must regularly monitor and test the operation, efficacy, and risk of individual AI applications, and mitigate emerging risks identified through monitoring, they are not explicitly required to monitor for developing bias and algorithmic discrimination. Perhaps that is implied by the earlier command to "[i]dentify and mitigate factors that may contribute to unlawful discrimination or harmful bias, including through determining whether the AI model results in significant disparities in the model's performance . . . across demographic groups,"[253] or perhaps that earlier command only applies to pre-deployment testing. It seems problematic that in national security settings, where the stakes for rights-impacting AI are often the highest, the requirements for monitoring for discriminatory bias are less rigorously drafted, leaving room for them to be less rigorously applied.

Accountability. Like the 2020 Intelligence Community Ethical Framework for AI, these 2024 executive policy documents collectively show a sophisticated understanding of the risks of algorithmic discrimination and the need to constantly monitor AI applications for bias throughout their lifecycles. But even at their strongest, as in the OMB Memorandum M-24-10, the policies allow for considerable discretion by agency employees. Agency officials evaluate, test, and monitor AI applications, then ultimately determine whether to use it. Of course, officials *should* do all those things, and such discretion allows room for necessary nuance, given that every AI application is different. But that same discretion also allows room for mistakes and abuse.

For national security systems falling outside the OMB framework, agency officials will have equal or greater discretion, given that their policy choices will be exercised in a classified environment. As argued above, remedies will be harder to come by in court or otherwise. The discretion and secrecy afforded to national security agencies suggests that even greater oversight is required for national security AI systems.

The NS AI Framework makes some progress toward accountability by providing for Chief AI Officers and Governance Boards, and outlining issues they and other officials, including privacy and civil liberties officials, should undertake to develop further oversight measures.[254] On accountability, the NS AI Framework requires that agencies will, among other things, "establish appropriate mechanisms to hold relevant personnel, including AI developers,

252. *Id.*
253. NS AI Framework, *supra* note 226, at 6.
254. NS AI Framework, *supra* note 226, at 9–12.

operators, and users, accountable for their contributions to and use of AI system decisions and action."[255] It would be significant to add to that requirement an explicit statement that high-level agency officials shall be held accountable for agency decisions to adopt and continue to use particular AI applications. Importantly, oversight must also be exercised from outside the Executive Branch, by Congress and courts, as outlined in the recommendations in Part V below. That is especially true if AI regulation is to withstand changes in Presidential administrations.

Public Transparency. Additionally, if AI used for national security purposes is to have public legitimacy, at least some transparency is necessary. Both the OMB and NS AI Frameworks contemplate some increased transparency, through the publication of AI use cases by agencies covered by the OMB memorandum,[256] and of *categories* of prohibited and "high impact" uses of AI covered by the NS AI Framework, though classified annexes are permitted.[257] Individual national security use cases are not required to be published, nor are categories of use cases below the "high impact" threshold.[258]

While concerns about protecting sources and methods will preclude the possibility of many national security use cases being posted publicly to the internet, agencies should lean toward providing general disclosures of both categories and individual AI use cases where possible. Likewise, the more that the executive branch can continue to publish unclassified regulatory documents, such as the NS AI Framework, or even the general contours of further regulations, the better. The public does not know the details of the FISA 702 program, for example, but due to work by the Privacy and Civil Liberties Oversight Board and disclosures by the Office of the Director of National Intelligence and the National Security Agency, we have a much better sense of the program's workings than when it was first deployed.[259] Likewise, the guiding regulations and general uses of national security AI should be disclosed whenever possible, consistent with national security.

255. NS AI Framework, *supra* note 226, at 14.

256. OMB Memorandum M-24-10, *supra* note 223, § 3(a).

257. NS AI Framework, *supra* note 226, at 11.

258. *See id.* at 11–12.

259. *See, e.g.*, reports and publications explaining FISA 702 procedures on those agency's websites: PCLOB: https://www.pclob.gov/Oversight; ODNI: https://www.dni.gov/index.php/newsroom/reports-publications/reports-publications-2023/3672-fisa-section-702-resources; and NSA: https://www.nsa.gov/Signals-Intelligence/FISA/.

V. RECOMMENDATIONS TO MITIGATE AND CHALLENGE DISCRIMINATORY AI

Governing AI in the national security space will be a wicked problem, especially with respect to the inevitable bias and algorithmic discrimination that will emerge. Keeping in mind that in many instances, the best choice may be to forgo using AI altogether for particular purposes, here are some recommendations for Congress, executive branch attorneys and policymakers, and courts and civil rights advocates.

A. OVERARCHING RECOMMENDATIONS

Legal scholarship on AI governance suggests that we must use both ex ante and ex post regulatory tools.[260] Ex ante tools help provide preventative and systemic protections against AI harms, some of which, as Professors Gianclaudio Malgieri and Frank Pasquale write, are "too serious to be recompensed ex post."[261] On the front end, I recommend that all government actors should be required to retain records relating to AI data, algorithms, design choice, and performance tests pre-and post-deployment;[262] to perform detailed risk assessments[263] and consult meaningfully with stakeholders, as contemplated in OMB Memorandum-M-24-10, described above; to test for bias before deployment and routinely thereafter; to develop bias monitoring and reporting systems; and to cease use of discriminatory AI. Ideally, these ex ante requirements should be imposed by legislation rather than by executive order. I also argue, below, for Congress to create a court to review any classified government AI models for bias before and while they are implemented.

Ex post tools address individual harms, and as Professor Margot Kaminski argues, provide a feedback mechanism for government actors to better

260. *See* Margot E. Kaminski, *Regulating the Risks of AI*, 103 B.U. L. REV. 1347, 1367–80 (2023); Gianclaudio Malgieri & Frank Pasquale, *Licensing High-Risk Artificial Intelligence: Toward Ex Ante Justification for a Disruptive Technology*, 52 COMPUTER L. & SEC. REV. 105899 (2024) (arguing for "an ex ante justification model to complement ex post regimes" to protect individuals and communities against harms, and calling for ex ante licensing requirements for private actors to demonstrate an AI tool's security, non-discrimination, accuracy, appropriateness, and correctability before it is deployed).

261. Malgieri & Pasquale, *supra* note 260, at 3.

262. *See* Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan & Cass R. Sunstein, *Discrimination in the Age of Algorithms*, 10 J. OF LEGAL ANALYSIS 113, 114 (2018) (noting the importance of a requirement that "all of the components of an algorithm (including that training data) . . . be stored and made available for experimentation[.]").

263. Pasquale & Malgieri, Kaminski, and others have written about risk impact assessments. Kaminski, *supra* note 260, at 1368.

regulate AI harms on the front end.[264] With respect to ex post tools and approaches, I recommend that Congress provide for several types of causes of action for discriminatory AI, described below. Courts and litigants should treat algorithmic bias as objectively measurable and therefore remediable, even under existing legal precedent.

B. FOR CONGRESS

1. Provide for damages relief for AI disparate treatment.

Congress should provide a statutory cause of action for damages relief for discriminatory AI. Although it would be preferable for the Court to provide a *Bivens* remedy where Constitutional rights have been violated, recent precedent suggests that is a dead end. Congress, therefore, should take the *Ziglar*[265] and *Egbert*[266] Courts up on the suggestion that Congress provide statutory causes of actions for constitutional violations by federal officials, as it has done with respect to state officials in Section 1983 claims. Congress might do so with respect to constitutional torts generally, or constitutional torts in the national security context, or specifically with respect to violations caused by or exacerbated by discriminatory AI, perhaps in the context of a larger AI bill.[267]

In addition to providing a statutory damages remedy for disparate treatment by AI, Congress might also establish statutory causes of action for damages for First, Fourth, and Fifth Amendment violations by government use of AI.

2. Provide a statutory basis to bring disparate impact suits for both injunctive and damages relief.

Though I have argued above that the use of discriminatory AI amounts to disparate treatment, Congress could certainly make it easier for civil rights advocates by providing disparate impact causes of action. Congress should establish causes of action for both injunctive and damages relief in disparate impact suits against federal, state, and local government actors using discriminatory AI.[268] The Court has suggested that Congress has the authority to do so in *Washington v. Davis*, where the Court opined that any extension of

264. Kaminski, *supra* note 260, at 1397.

265. Ziglar v. Abbasi, 582 U.S. at 135–36 ("The proper balance is one for the Congress, not the Judiciary, to undertake.").

266. Egbert v. Boule, 596 U.S. at 502.

267. If Congress took that last approach, it might caveat that in providing for a specific damages remedy with respect to algorithmic discrimination, it did not intend to preclude the courts from remedying constitutional torts in other contexts.

268. Congress might also consider authorizing disparate impact suits against private actors, though the Court might be less likely to uphold it.

a rule allowing disparate impact suits "should await legislative prescription."[269] As Dean Erwin Chemerinsky writes, "civil rights statutes can, and often do, allow violations to be proved based on discriminatory impact without evidence of a discriminatory purpose."[270] Congress has authorized disparate impact suits under Title VII of the 1964 Civil Rights Act for employment discrimination and the 1982 Amendments to the Voting Rights Act of 1965.[271]

In providing either monetary damages for disparate treatment by AI (as suggested above), or injunctive or monetary for disparate impact due to AI, Congress might implicitly, or explicitly, express its will that using biased AI violates equal protection law. Establishing that using biased AI is illegal will help prevent the use of qualified immunity defense that the law was not "clearly established."

3. Establish an oversight court for classified AI systems that affect civil rights and civil liberties.

Given how many barriers exist to keep constitutional challenges in the national security context from being heard on the merits, Congress should provide for oversight before AI is ever used. Ideally, this would be a court dedicated to constitutional rights review of classified government AI systems.[272] Congress might establish it as an Article III court, like the U.S. Foreign Intelligence Surveillance, or an Article I court. Or, failing Congressional action, the executive branch might create, by executive order, a court within the executive branch, like the Data Protection Review Court established in 2022.[273]

---

269. Washington v. Davis, 426 U.S. 229, 248 (1976).

270. CHEMERINSKY, *supra* note 141, at 770.

271. *Id.*

272. Professor Ashley Deeks has an interesting proposal worth exploring that there should be a covert action-like congressional approval system for AI. Ashley Deeks, *Regulating National Security AI Like Covert Action?*, LAWFARE BLOG (July 25, 2023), https://www.lawfaremedia.org/article/regulating-national-security-ai-like-covert-action. Her focus seems to be largely on safety and international repercussions, traditional covert action concerns. I would add to that proposal, if it is not already implicitly included, that any direct oversight by congressional committees might also include briefing on disparate impact, disparate treatment, and other related harms. That civil rights and civil liberties review for discriminatory bias should only compliment, not preclude, a dedicated court.

273. *See* U.S. Dep't of Justice, Data Protection Review Court, 28 C.F.R. 201, 87 Fed. Reg. 62303 (Oct. 14, 2022); Peter Margulies, *Adjudicating Algorithms: Accountability in Regulation of Surveillance, Privacy, and Discrimination*, Roger Williams Univ. Legal Studies Paper No. 216, 28–33 (Mar. 1, 2023) (working paper) (discussing key features of the Data Protection Review Court, independence and the appointment of a Special Advocate, but also the limits of a court within the executive branch and with judges without lifetime tenure). Professor Margulies has suggested an "Algorithmic Rights Court," with broader jurisdiction over various types of AI

Before using any classified AI system, the government would have to present that system to the court for review, including metrics about its training, testing, and validation data; its inputs, weights, and outputs; and assessments of any disparate treatment or impact. The court would look specifically for equal protection, First and Fourth Amendment, and due process concerns for each classified AI model, with a special focus on bias. Agencies would be required to file for approval before implementing an AI system or model. Agencies would also be required to file updates at regular intervals during the AI's deployment to provide monitoring data. As part of these filings, agencies would need to produce (and therefore preserve)[274] all records related to AI design decisions; training, testing, and validation data; and performance metrics on bias.

C. FOR EXECUTIVE BRANCH ATTORNEYS

As ever, government attorneys in classified environments must be guardians of constitutional rights in situations where the courts and Congress have yet to intervene. Government attorneys should learn the fundamentals of AI, so that they might talk with technologists throughout the lifecycle of the AI, from inception to end of use. They should ask pointed questions at every step to test for and weed out bias. Attorneys should insist on understanding the inputs, weights and outputs of every algorithm; in particular, they should know what factors might place an individual under heightened suspicion or surveillance by the government. Those factors should never include suspect categories or the many possible proxies for suspect categories, including innocent social behaviors, but rather only concrete, individually based behaviors directly correlated with an anticipated criminal or foreign intelligence/counterintelligence activity, such as buying necessary technology or equipment. Agencies should establish procedures for regular monitoring of their AI applications for bias. Agencies should also establish internal oversight measures, using IG offices, Chief AI Officers and Governance Boards, and Civil Rights and Civil Liberties Officers and offices to aid in this bias monitoring process. Intelligence Oversight Board reporting mechanisms should be updated to include questions regarding whether attorneys and officials have seen or are concerned about instances of algorithmic discrimination and other rights or safety issues. To any government attorneys or officers reading this: thank you for doing this difficult, detailed, and unseen work.

---

harms, including databreaches and cybersecurity concerns, than the court proposed in this Article. *See id.* at 34.

274. *See* Kleinberg et al., *supra* note 262, at 114.

D. FOR COURTS AND LITIGANTS

To summarize my more detailed recommendations developed in Part III, both courts and civil rights litigants should treat AI as objectively measurable, though not objective, and therefore reviewable by courts in multiple contexts.

Regular Fourth Amendment concerns about reasonableness, and not *Whren*, should apply to discriminatory AI.

Under equal protection doctrine, any government use of biased AI should be treated as a facially discriminatory policy; the AI's outputs should be assessed at the moment it is to be deployed, or is being deployed, to see whether the AI is classifying people by suspect categories. In the alternative, at the very least, the government's choice to use discriminatory AI should be treated as purposeful disparate treatment, and strict scrutiny should apply.

Due process demands transparent, trustworthy, and fair AI. The AI must be transparent to someone—to the government throughout its development and use, and during litigation, and to the court and cleared counsel, at a minimum. If the government cannot demonstrate or refuses to demonstrate that the algorithms it uses provide equal, error-free treatment, courts should enjoin their use.

VI. CONCLUSION

Algorithmic bias should be treated under the law as what it is: exceedingly likely (perhaps inevitable) and objectively measurable. Algorithms that discriminate may do so in subtle or obvious ways, but in all cases, knowable, discoverable ways. Biased AI is facially discriminatory in that it separates and bins people along lines of suspect classification. Any choice to use it, especially by such sophisticated actors as the national security and law enforcement agencies, is purposeful discrimination. Government actors therefore have a responsibility to employ every methodology to reduce bias; such interventions should be viewed as anticlassification tools rather than affirmative action. Where AI is used, it must be tagged to specific behaviors that in themselves might constitute part of criminal activity, such as purchasing equipment necessary for a criminal act, rather than to any social identity descriptors or social behaviors aligned or correlated with suspect classifications. If discriminatory bias cannot be eliminated, the government has a constitutional obligation not to use that AI. Such a choice would also be good security policy as it would avoid inaccurate intelligence or investigatory conclusions.