AI, Bias, and National Security Profiling

Laurie Hobart[1]

Abstract:

Artificial Intelligence, 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 on the basis of ethnicity and 'race', national origin, and religion. AI is or may be used in national security criminal investigations; intelligence, counterintelligence, or counterterrorism activities; watch listing practices; border policies and customs investigations; and general monitoring or surveillance programs.

This article outlines the ways that AI may exacerbate and reproduce at scale existing bias in national security investigations. It 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. Part I briefly explains from a technical perspective how bias is produced and reproduced by AI, how bias might alter investigatory and intelligence outcomes, and how AI might expand the net of people under surveillance. Part II details how existing national security case law in combination – including Fourth Amendment, Equal Protection, First Amendment, and Due Process precedents; limitations on Bivens claims; and evidentiary privileges such as executive privilege, classification generally, and state secrets doctrine – is inadequate and even problematic for the protection of civil rights and liberties from harmful uses of AI. While detailing the ways in which current case law is not 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. For example, I argue why 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. I also argue why, under equal protection law, AI bias should be treated as actionable disparate treatment rather than non-actionable disparate impact. Part III addresses executive policies, such as the Department of Justice "Guidance for Federal Law Enforcement Agencies Regarding the Use of Race, Ethnicity, Gender, National Origin, Religion, Sexual Orientation, Gender Identity, and Disability," the Intelligence Community's Artificial Intelligence Principles and Ethics Framework, and the October 2023 executive order on AI. Those policies have both problematic and helpful aspects. Part IV suggests solutions civil rights and civil liberties advocates might pursue in litigation and legislation. It also provides recommendations for courts and government attorneys seeking to minimize bias or national security profiling by humans and machines.

[1] Associate Teaching Professor, Syracuse University College of Law; former Assistant General Counsel, Central Intelligence Agency.

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 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 Latin American immigrants, culminating in the border wall and the separation of families and, again, the taking of children; the over-policing of Latinix (Latino) Americans; the rejection and refoulement of asylum applicants; the round 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 immigrants 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, autocratic measures and bigotry advocated for openly by politicians at home, and another presidential election soon upon us, 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 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 tracking we submit to by the millisecond and phone swipe by any number of private companies, government agencies, and rogue internet actors. AI tools are being employed in all fields: medicine and health care, website 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 create solutions, perhaps to climate change, or horrors, such as biological weapons.[2] AI has the potential to bring great benefits to humanity but also carries great risks.[3] 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.[4]

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

[3] *See id.; Biden, Harris meet with CEOs about AI risks*, AP NEWS (May 4, 2023), https://apnews.com/article/ai-artificial-intelligence-white-house-harris-578d623e473b0eeb3fa3e4728d7e9868

[4] 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 Assessment Tools in the Criminal Justice System* (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 Technical Flaws of Pretrial Risk Assessments Raise Grave Concerns*, BERKMAN KLEIN CENTER FOR INTERNET & SOCIETY AT HARVARD UNIVERSITY (Jul. 17, 2019), https://cyber.harvard.edu/story/2019-07/technical-flaws-pretrial-risk-assessments-raise-grave-concerns

The algorithms employed are biased – often racist, perhaps also gendered, classist, or otherwise unfair – and produce biased results. Scholarly and media attention there is critical. This article, however, focuses on the potential for bias and for AI profiling by different 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 on the basis of ethnicity and 'race', national origin, and religion, and to reproduce that bias at scale. AI is or might 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. The government has a pressing need to use AI for at least some national security purposes– that is well argued and documented.[5] But the legal guardrails are shaky, and at some points along the highway, missing altogether. The pace of both national security practice and AI development are very fast, so guardrails are especially needed. As is often observed, the law has not kept pace with technological development. While AI is inherently risky for civil rights and civil liberties, and perhaps unsolvably so, it may well be the case that current case law poses equal or even greater threat to our constitutional values.

Here, I outline the ways that AI might exacerbate and reproduce at scale existing bias in national security investigations and surveillance; argue 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 suggest policy solutions for Congress and executive agencies, and litigation strategies for plaintiffs. Part I will briefly explain from a technical perspective how bias in produced in and reproduced by AI, how bias might alter investigatory and intelligence outcomes for the worse, and how AI might expand the net of people under surveillance. Part II details how existing national security case law in combination – including Fourth Amendment, Equal Protection, First Amendment, and Due Process precedents; limitations on *Bivens* claims; and evidentiary privileges such as executive privilege, classification generally, and state secrets doctrine – is inadequate and even problematic for the protection of civil rights and liberties against harmful uses of AI. While detailing the ways in which current case law is not 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. For example, I argue that the Supreme Court case *Whren v. United States,* 517 U.S. 806 (1996), 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 Equal Protection law. Part III discusses recent executive policies, such as the Department of Justice "Guidance for Federal Law Enforcement Agencies Regarding the Use of Race, Ethnicity, Gender, National Origin, Religion, Sexual Orientation, Gender Identity, and Disability," the Intelligence Community's Artificial Intelligence Principles and Ethics Framework, and the October 2023

[5] *See* NATIONAL SECURITY COMMISSION ON ARTIFICIAL INTELLIGENCE, FINAL REPORT, Ch. 5 (2021), https://reports.nscai.gov/final-report/ ; Corin Stone, *The Integration of Artificial Intelligence in the Intelligence Community: Necessary Steps to Scale Efforts and Speed Progress,* JOINT PIJIP/TLS RESEARCH PAPER SERIES (2021) https://digitalcommons.wcl.american.edu/research/73

executive order on AI. These policies have both problematic and helpful aspects. Part IV suggests solutions civil rights and civil liberties advocates might pursue in litigation and legislation. It also provides recommendations for government attorneys seeking to minimize bias or national security profiling by humans and machines.

I. How Bias Is Produced in AI and Reproduced at Scale

Part I briefly explains from a technical perspective how bias is produced and reproduced by AI, how bias might alter investigatory and intelligence outcomes, 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 variance between the desired accuracy of an AI model and the actual output. This variance might include, hypothetically, an AI designed to recognize an enemy tank at night, trained on images of tanks at night, that when actually deployed mistakes, in the dark, the box like shape of an air conditioner on the side of a building for a tank.[6] There is 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. 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."[7]

Algorithmic bias includes discriminatory bias, where the AI model disproportionately affects or discriminates against a group of people, such as on a racial, ethnic, gender, religious, nationality, disability, or sexuality-based classification. Discriminatory bias might determine, for example, who is being delayed or detained at the airport or border. The Biden Administration's recent executive order on AI and the Draft AI Bill of Rights[8] helpfully use the term "algorithmic discrimination" when referring to this type of bias in AI.

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

[6] James E. Baker, Laurie Hobart, Matthew Mittelsteadt, and John Cherry, National Security Law and the Coming AI Revolution: Observations from a Symposium Hosted by Syracuse University Institute for Security Policy and Law and Georgetown Center for Security and Emerging Technology, Oct. 29, 2020 (with Jamie Baker, Matt Mittelsteadt, and John Cherry) (2021), https://cset.georgetown.edu/wp-content/uploads/Symposium-ReportNational-Security-Law-and-the-Coming-AI-Revolution.pdf

[7] KUSH R. VARSHNEY, TRUSTWORTHY MACHINE LEARNING 16 (2022), http://www.trustworthymachinelearning.com/ Dr. Varshney leads the machine learning group in the Foundations of Trustworthy AI department at IBM and co-directs the IBM Science for Social Good initiative.

[8] Draft AI Bill of Rights ; Executive Order on AI

Bias can enter an AI model in myriad ways. Here, I will discuss some of the most common ways, but this is not an exhaustive list or treatment.[9] This discussion will follow the stages of the machine learning[10] lifecyle – that is, the lifecycle of an AI model.[11] Lawyers need to know that human judgement, values, and biases are inserted into AI models at every step of that process. Moreover, every AI model is unique, and not only that, evolving and learning (for better or for worse) over the course of its "life," from its conception and design through its training and deployment in the field through its dying days (end of use).

Bias can be produced by the original framing of the question or task posed to the AI to answer or perform.[12] Designing a facial recognition system for unlocking a phone has a different ethical implication from creating one to identify and track a particular population, such as the Uighur people in China. (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.

Likewise, human values and biases inform the metrics by which we measure the AI's accuracy or success[13]: 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? *Etc.* One can build as many (or 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[14] (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.[15] Representation bias can also include differences in quality or labeling or engineering of data by technologist across groups.[16] Algorithms used for predictive policing and for 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 on which any AI would be trained reflect the *historically disproportionate* policing, arrest, prosecution, and sentencing of people of color for the same conduct as white

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

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

[11] Varshney, *supra* note 19, at 14–22 (citing the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology); *see also* Nick Holtz, *What is the Data Science Process?*, DATA SCI. PROCESS ALL., https://www.datascience-pm.com/data-science-process/ (last visited Sept. 7, 2023) (identifying the CRISP-DM and other life cycle models for data science).

[12] Varshney, *supra* note __ at __.

[13] *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).

[14] *See* Varshney, *supra* note __ at 48.

[15] *Id.* at 48.

[16] *Id.*

people.[17] That is an example of "population bias," where (data about) people of color are overrepresented in the training and possibly testing and validation data, and white people are underrepresented. Another example of population bias is that early facial recognition algorithms performed especially poorly at accurately matching faces of women of color because similar faces were underrepresented in the training data.[18]

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, such as zip codes for demographic information (racial, ethnic, etc.) or, in some cultures, last names for religion.[19] Housing and employment data used in criminal risk assessments have proved to be strong proxies for "race" and class.[20] 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."[21] Feature engineering data involves "mathematically transforming features to derive new features . . . . Apart from the initial problem specification, feature engineering is the point in the [machine learning] lifecyle that requires the most creativity from data scientists."[22]

Once the data is prepared, engineers will have to 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 will select the inputs – what features, expressed in numbers, the model will evaluate or process, as well as all of the parameters or connections between inputs and later interpretations of the inputs, as well as 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.[23] This lack of transparency poses significant due process and equal protection issues. There are, however, other methodologies available: there is increasing research into more transparent neural networks that will allow more insight into the decisions, weights, and parameters.[24]

A model is evaluated or tested on a separate, held-out set of "validation" data.[25] That is, it is fed inputs from that data set to see whether it works and how well. Validation data can also be a source of bias as the AI learns and potentially readjusts its algorithms based on that data.

[17] *Technical Flaws of Pretrial Risk Assessments Raise Grave Concerns*, BERKMAN KLEIN CENTER FOR INTERNET & SOCIETY AT HARVARD UNIVERSITY (Jul. 17, 2019), https://cyber.harvard.edu/story/2019-07/technical-flaws-pretrial-risk-assessments-raise-grave-concerns [supra note

[18]

[19] See Varshney, *supra* note __, at 18.

[20] Chelsea Barabas et al., Open Letter to the Members of the Massachusetts Legislature Regarding the Adoption of Actuarial Risk Assessment Tools in the Criminal Justice System (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)).

[21] Varshney, *supra* note 5, at 19.

[22] *Id.*

[23] *Id.* at 21.

[24] *See id.* at 21(?); *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.

[25] Varshney at 21.

Finally, the AI is deployed 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.[26] 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.[27]

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

Although I will make arguments below why *the choice to use* biased AI may be considered an intentionally discriminatory act under current case law, the human bias that is built into the AI might be explicit or implicit, intentional or unconscious. It is objectively measurable, but it may or may not have been *consciously* intended by the designers. As Professor Charles Lawrence argued in 1987,

> Americans share a common historical and cultural heritage in which racism has played and still plays a dominant role. Because of this shared experience, we also inevitably share many ideas, attitudes, and beliefs that attach significance to an individual's race and induce negative feelings and opinions about nonwhites. To the extent that this cultural belief system has influenced all of us, we are all racists. At the same time, most of us are unaware of our racism. We do not recognize the ways in which our cultural experience has influenced our beliefs about race or the occasions on which those beliefs affect our actions. In other words, a large part of the behavior that produces racial discrimination is influenced by unconscious racial motivation.[28]

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.[29] Since that writing, there has much research on social and cognitive biases,[30] all important to understanding how we may program bias into AI. Of central importance to Professor Lawrence, he later wrote, however, was "the cultural meaning of racial texts:"[31] he sought to explore "how white supremacy is maintained not only through the intentional deployment of

[26] Id.

[27] Oscar Schartz, *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.

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

[29] *Id.* at 322.

[30] *See* Charles Lawrence, *Unconscious Racism Revisited: Reflections on the Impact and Origins of "The Id, the Ego, and Equal Protection,"* 40 CONN. L. REV. 931-978 (2008).

[31] *Id.* at 938.

coercive power, but also through the creation, interpretation, and assimilation of racial text."[32] When humans program AI, machines will read our cultural texts, with all of their biases and prejudices, 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.[33] Historically, the intelligence and law enforcement communities targeted the civil rights movement, the anti-war movement, Muslims, and more, because of perceived "threats." One of the reasons it is so important to diversify the Intelligence Community 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. [[Add material here, including on point of view and positionality]]. But that ongoing[34] work is far from complete,[35] and current biases may be reflected in the development of or choice to use particular AI models or AI for particular purposes.

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."[36] With AI, we have a collective duty to counter the collective cultural racisms and other biases that it will invariably collect and disseminate, and to redress its 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. Due to classification, to some extent we really must imagine the risks and harms, but here 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 watch listing, for example, where the U.S. government has historically disproportionately included Muslims and Muslim Americans.[37] Or AI might be

[32] *Id.* at 939.

[33] See Shirin Sinnar, *Separate and Unequal: The Law of "Domestic" and "International" Terrorism*, 117 Mich. L. Rev. 1333, 1388 n. 307 (2019).

[34] See, e.g., The Intelligence Community Centers for Academic Excellence (IC CAE) Program, https://www.dni.gov/index.php/iccae; and the ODNI's page on Diversity and Inclusion, https://www.dni.gov/index.php/how-we-work/diversity.

[35] See, e.g., GAO-21-83, *Intelligence Community: Additional Actions Needed to Strengthen Workforce Diversity Planning and Oversight* (Dec. 17, 2020), https://www.gao.gov/products/gao-21-83.

[36] Charles Lawrence, *Unconscious Racism Revisited*, *supra* note ___, at 942.

[37] *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.").

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.[38] For example, it often categorizes threats of terrorism by Muslim Americans to be 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.[39] This false dichotomy poses significant harms to U.S. Muslim individuals, such as harsher sentencing, and communities, such as disproportionate surveillance[40] and "distorted public perceptions of terrorism that fuel anti-immigrant and discriminatory policies."[41] Professor Sinnar also suggests that the category of "international" terrorism "will predictably expand to cover U.S. individuals perceived as 'foreign,' even if they are citizens with negligible relationships abroad."[42] Significantly, the false dichotomy may cause us to underestimate the risk of violent harm by "domestic" terrorists.[43] AI that adopts the dichotomy will increase all of those harms; and it will be inaccurate.

Moreover, any integrated government systems reliant on the outputs produced by the AI will also be infected by it. The potential for harm – and biased harm – is evidenced by analogy in watchlisting cases, even without AI involved. In *Ibrahim*,[44] one government employee's mistake infected multiple watch lists. In *Elhady*,[45] the Eastern District of Virginia, though later reversed on other grounds, determined that the central national database from which all other, shorter lists are derived, the Terrorist Screening Database (TSDB), posed due process issues, including the low standard – the executive's reasonable suspicion standard – for inclusion on the lists.

2. Government Aggregation of Data Types and Systems

If various government agencies and private entities merge and data mine the various types of data they have collected, greater privacy invasion and greater harms from errors will result.[46] Data collections might be integrated across types (biometric information with tax information with financial records with criminal records, etc.) and across collecting entities: law enforcement with intelligence, federal with state and local, etc. 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

[38] Shirin Sinnar, *Separate and Unequal: The Law of "Domestic" and "International" Terrorism*, 117 Mich. L. Rev. 1333 (2019).
[39] *Id.* at 1337
[40] *Id.* at __
[41] *Id.* at __
[42] *Id.* at __
[43] *See id.* at 1388-92
[44] *Ibrahim v. Department of Homeland Sec.,* 62 F. Supp. 3d 909 (N.D. Cal. 2014).
[45] *Elhady v. Kable*, 391 F. Supp. 3d 562 (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/.
[46] *See* Stephen Dycus, Williams C. Banks, Peter Raven-Hansen, and Stephen I. Vladeck, NATIONAL SECURITY LAW, 7th. ed., 739 (2020), citing Daniel J. Solove, *A Taxonomy of Privacy*, 154 U. Pa. L. Rev. 477, 507 (2006).

law enforcement organizations, and other public and private entities."[47] 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."[48] Data aggregation and harvesting, of course, are or will 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 that will infiltrate and reproduce in linked AI datasets and systems? It is 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[49] legislation. there is nothing to protect against the privacy invasion of aggregated data except the application of the Fourth Amendment. Like Congress, the courts have been reticent to articulate the ground rules. As discussed below, the 2018 case *Carpenter*, where the Supreme Court held that police needed a warrant to obtain 127 days-worth of cell cite data, provides some hint that it 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 increase the net of surveillance thrown over Americans and people around the globe. AI enables, among other things, facial recognition and other biometric analysis, remote cameras and drones, data aggregation, data mining, link analysis, *etc.* 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 v. United States*.[50] AI-enabled surveillance occurs or may occur not only in public spaces but also in our homes and offices,[51] via our computers, phones, personal electronic assistants, wearable heath monitors, connected appliances and cars, and home 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.[52] Such surveillance potential creates

[47] Margaret Hu, *Biometric Cyberintelligence and the Posse Comitatus Act*, 66 Emory L. J. 697, 701 (2017).
[48] Hu, *supra* note 38, at 700-01.
[49] Note Illinois and other states
[50] *Carpenter*, 1138 S. Ct. at 2210. *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.").
[51] There is a growing body of literature on AI in the workplace. *See, e.g.*, Karen E. C. Levy (2015), *The Contexts of Control: Information, Power, and Truck-Driving Work, The Information Society*, 31:2, 160–74, https://www.tandfonline.com/doi/full/10.1080/01972243.2015.998105 (accessed Oct. 27, 2019).
[52]

tremendous First and Fourth Amendment problems, among other civil and human rights issues, the subject of much scholarship.[53]

Of relevance here, AI increases the scope of government investigations, enabling surveillance of many more people than traditional gum shoe police work.[54] Any discriminatory bias will likewise have expansive effect, harming more people. Facial recognition systems offer an example of this expanding ripple effect. Many systems, including the FBI's, 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.[55] Facial recognition might flag someone who was five states away from the crime scene; as Jennifer Lynch argues, in making that person 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. This is true even if a face recognition system offers several results for a search instead of one; each of the people identified could be brought in for questioning, even if there is nothing else linking them to the crime. Former German Federal Data Protection Commissioner Peter Schaar has noted that false positives in face recognition systems pose a large problem for democratic societies: "[I]n the event of a genuine hunt, [they] render innocent people suspects for a time, create a need for justification on their part and make further checks by the authorities unavoidable.[56]

Moreover, she suggests, biased AI models and databases mean that harm of false positives will be disproportionately felt by people of color.[57]

[53] *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*, ELECTRONIC FRONTIER FOUNDATION 1, 8–10 (Feb. 12, 2018), https://www.eff.org/wp/law-enforcement-use-face-recognition#_idTextAnchor004; Claire Garvie, Alvaro M. Bedoya & Jonathan Frankle's *The Perpetual Lineup: Unregulated Police Face Recognition in America*, Georgetown Law Center on Privacy & Technology (Oct. 16, 2016), https://www.perpetuallineup.org/.
53

[54] 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* (2012)("In the pre-computer age, the greatest protections of privacy were neither constitutional nor statutory, but practical. Traditionally surveillance for any extended period of time was difficult and costly and therefore rarely undertaken. The surveillance in this case – constant monitoring of the location of a vehicle for four weeks – would have required a large team of agents, multiple vehicles, and perhaps aerial assistance.")

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

[56] *Id.*

[57] *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.")

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 classification, those groups will feel the harm of false positives disproportionately. If, for example, an AI is used to predict who is most likely to commit a particular national security-related criminal act, a biased AI might falsely accuse members of those groups disproportionately 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,[58] if the government has as 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 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 any of those people is individually likely to be 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, on the basis of 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 Surveillance

The October 30, 2023 Executive Order on AI[59] provides "requirements" for government use of AI *other than in national security systems*: non-national security agencies should, "where appropriate," adopt the following practices:

> "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."[60]

Some published uses of AI by the government include: [ADD examples from State, DHS, etc. websites].

Classified systems, however, are generally unknown.[61] The Executive Order simply provides that the Assistant to the President for National Security Affairs and the Assistant to the

[58] U.S. Department of Justice, "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/2023-06/Guidance%20for%20Federal%20LEAs%20on%20the%20Use%20of%20Protected%20Characteristics_FINAL%205.25.23_508.pdf

[59] E.0. 14110 on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (Oct. 30, 2023), https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence//

[60] Id. at 10.1(b)(iv).

[61] [Add any research from unclassified sources.]

President and Deputy Chief of Staff for Policy "shall oversee an interagency process with the purpose of, within 270 days of the date of this order, developing and submitting a proposed National Security Memorandum on AI to the President."[62]

One way to think about the potential for classified AI surveillance is to consider that the government could, with only the Constitution standing in its way (i.e. no 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 or the expectation of legal accountability for violations. Government officers – lawyers and the policy officials they advise – typically practice in secrecy. This 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 view, ethically unsound legal advice where constitutional rights stand to be violated. As the Courts will comment,[63] just because 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.

II. 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 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.[64] I will briefly discuss how AI will only make this 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**

[62] E.0. 14110, supra note __, at 4.8.

[63]

[64] Dycus, Banks, Raven-Hansen, and Vladeck, *supra* note __, chapter 5.

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:

> [T]he Constitution prohibits selective enforcement of the law based on considerations such as race. But the constitutional basis for objecting to intentionally discriminatory application of laws is the Equal Protection Clause, not the Fourth Amendment. Subjective intentions play no role in ordinary, probable-cause Fourth Amendment analysis.[65]

In *Whren* and subsequent cases, the Court held that "the subjective intent of the officer is irrelevant" for Fourth Amendment purposes,[66] 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. 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. [67]

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"; "I stopped too quickly at the stop sign"; jaywalking, etc.).[68]

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.[69] 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.[70]

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") but criminal justice advocates question how easy or likely it is for profiled persons to file, much less win, equal

[65] 517 U.S. 806, 813 (1996).

[66] *Florida v. Jardines*, 569 U.S. 1, 10 (2013)

[67] *Florida v. Jardines*, 569 U.S. 1, 10 (2013)

[68] Cite 5-4 Podcast and personal experience with client

[69] Cite 5-4 Podcast

[70] See Plaintiff's Brief in Opposition to City's Motion to Dismiss, at footnote 3, https://ccrjustice.org/sites/default/files/assets/Hassan%20-%20MTD%20Response%20Final%201%2025%2013.pdf

protection claims. As discussed in the next section, 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 *Whren*:

> The Court's jurisprudence, I am concerned, sets the balance too heavily in favor of police unaccountability to the detriment of Fourth Amendment protection. A number of commentators have criticized the path we charted in *Whren v. United States* [. . .] and follow-on opinions, holding that "an arresting officer's state of mind . . . is irrelevant to the existence of probable cause," *Devenpeck v. Alford*, 543 U. S. 146, 153, 125 S. Ct. 588, 160 L. Ed. 2d 537 (2004). See, e.g., 1 W. LaFave, Search and Seizure §1.4(f), p. 186 (5th ed. 2012) ("The apparent assumption of the Court in Whren that no significant problem of police arbitrariness can exist as to actions taken with probable cause, blinks at reality."). I 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.[71]

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 affect AI will have on the courts' application of *Whren* is uncertain. 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[72] 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* itself, suggests that where only profiling and no pretextual reason exists, *Whren* does not apply and the Fourth Amendment prohibits the search or seizure.

(1) 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 or unintentionally biased AI. Given some (but not all) AI models' lack of transparency – at least at

[71] *District of Columbia v. Wesby*, 138 S.Ct. 577, 593 (2018) (Ginsburg, J. concurring).

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

present[73] – 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, these arguments will be undercut by the fact that more transparent AI models are both available and being developed.)[74]

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 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.[75] 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.

Justice Scalia later wrote in *Ashcroft v. Kidd* that "the Fourth Amendment regulates conduct rather than thoughts."[76] An AI does not think, it calculates. Unlike the officer's thoughts, the AI's operating algorithm, data inputs, test data run and in-use performance measurements, etc. are all knowable, 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 the government employs and relies on AI is human "conduct" also subject to review.

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. *Whren* was not written with AI in mind. Moreover, the government should be consistently measuring the AI's inputs, outputs, 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

[73] 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. Cite to AI for Judges and NIST study; *See also* Brandon L. Garrett and Cynthia Rudin, *Rethinking the Use of Artificial Intelligence in Criminal Justice,* forthcoming, 109 CORNELL L. REV. __ (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.")

[74] *See* sources in note 37.

[75] Cite Berkman Klein article on this; also, techchauvinism book?

[76] Ashcroft v. Al-Kidd, 563 U.S. ___ (2011) ("Fourth Amendment reasonableness "is predominantly an objective inquiry." Edmond, supra, at 47. We ask whether "the circumstances, viewed objectively, justify [the challenged] action." Scott v. United States, 436 U. S. 128, 138 (1978). If so, that action was reasonable "whatever the subjective intent" motivating the relevant officials. Whren v. United States, 517 U. S. 806, 814 (1996). This approach recognizes that the Fourth Amendment regulates conduct rather than thoughts, Bond v. United States, 529 U. S. 334, 338, n. 2 (2000); and it promotes evenhanded, uniform enforcement of the law, Devenpeck v. Alford, 543 U. S. 146, 153–154 (2004). "); 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.")

the fact 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; 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,"[77] and indeed the Court found a basis for probable cause independent of racism on the facts in that case.[78] 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 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. Just as but for subjective discriminatory bias on the part of a police officer patrolling the streets, a person of color might not have been pulled over for a traffic stop that escalates into more. 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 could certainly argue for throwing out any AI results altogether if the AI is shown to be infected with bias. If so infected, it will not provide the objective basis for probable cause that *Whren* requires.

One potential criticism of the argument advanced here, that review of AI under the Fourth Amendment should not be precluded by *Whren*, is that the Court's "primary concern" was not about the difficulty of proving that the police officer's subjective intent was racist or otherwise problematic, but rather whether other, objectively based probable cause existed. Justice Scalia wrote:

> If [precedent] cases were based only upon the evidentiary difficulty of establishing subjective intent, petitioners' attempt to root out subjective vices through objective means might make sense. But they were not based only upon that, or indeed even principally upon that. Their principal basis—which applies equally to attempts to reach subjective intent through ostensibly objective means—is simply that the Fourth Amendment's concern with "reasonableness" allows certain actions to be taken in certain circumstances, whatever the subjective intent.[79]

The evidentiary issue of measuring subjective intent was, however, one motivating concern. As referenced above, years later Scalia wrote that conduct rather than thoughts are reviewable under the Fourth. 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 reasonable actions. And therefore the AI, as argued above, becomes reviewable under the Fourth Amendment.

[77] See Whren at 812 or 813

[78] Whren at –last line

[79]

At a minimum, the government must prove that its AI is not simply and only making recommendations or predictions on the basis of suspect categories or proxies for those categories. As argued next, if there is no demonstrable and independent objective basis for probable cause, *Whren* does not preclude a review for reasonableness.

(2) 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, *Farag*,[80] dealt with that question and answered No, that where no legitimate reason 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 Amendment *once probable cause is established;* it was not called upon to address whether race might be relevant to the probable-cause analysis itself."[81]

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.[82] Surveying federal case law, the E.D.N.Y. district 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.

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." I would suggest, however, that even this use of 'race' might be problematic, for example, if law

[80] *Farag v. U.S.*, 587 F. Supp. 2d 436, 462-65 (E.D.N.Y. 2008)(" In *Whren,* the existence of probable cause based on non-racial factors was conceded. . . . . Thus, the Court opined only that an officer's subjective, potentially race-based motivations were irrelevant to the Fourth Amendment *once probable cause is established;* it was not called upon to address whether race might be relevant to the probable-cause analysis itself.)

[81] That is consistent with *Florida v. Jardines*, 569 U.S. 1, 10 (2013), which 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. ___ (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 concludes it was not.])

[82] *Farag v. U.S.*, 587 F. Supp. 2d 436, 443 (E.D.N.Y. 2008)

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, etc. "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."

This is the type of profiling infamously associated with the unwarranted arrest of Professor Henry Louise Gates, Jr., outside his own home in Cambridge, MA. 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 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 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."[83]

The *Farag* court rejected any use of "propensity" profiling as the basis for finding probable cause or contributing[84] to a finding of probable cause:

> Even granting that all of the participants in the 9/11 attacks were Arabs, and even assuming *arguendo* that a large proportion of would-be anti-American terrorists are 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.[85]

The court concluded that the bulk of precedent:

> . . . clearly evidences what has been described as an increasing "hostil[ity] to the use of race as a basis for police action under the Fourth Amendment." . . . . There is no doubt that the specter of 9/11 looms large over this case. Although this is the first post-9/11 case to address whether race may be used to establish criminal

[83] 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 at ---, quoting Brignoni-Ponce, 422 U.S. 873. 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."

[84]

[85] This is, notably, the opposite of the logic adopted by Justice Kennedy (for First and Fifth Amendment purposes) in *Iqbal*, discussed below.

> propensity under the Fourth Amendment, the Court cannot subscribe to the notion that in the wake of 9/11 this may now be permissible.[86]

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, if the biased AI is used to predict illegal behavior, that is likewise propsenity 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.

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 surveilles 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 stastically relevant the original suspect description becomes, because the likelihood that *any given* individual screened is the suspect becomes negligible.

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 based 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 even the current conservative 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 automous drones or other surveillance with video and sound sensors, the court might be even more worried about "near perfect surveillance,"[87] and require warrants, as some states have begun to do.[88] The First Amendment chilling effects of such surveillance are well established in scholarly literature.[89] But even *Carpenter* notably carved national security investigations out of its holding, declining to opine on them or require a warrant:

> We do not disturb the application of [third party doctrine cases] *Smith* and *Miller* or call into question conventional surveillance techniques and tools, such as security cameras. Nor do we address other business records that might incidentally reveal location information. Further, our opinion does not consider other collection techniques involving foreign affairs or national security.[90]

[86] *Farag v. U.S.*, 587 F. Supp. 2d 436, 467 (E.D.N.Y. 2008)
[87] Carpenter v. United States; AI and JUDGES FJC at -----;
[88]
[89] See, e.g, Margot Kaminski
[90] Carpenter v. United States at 18 in slip

Likewise, as so often is the case, 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.[91]

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 (requiring warrant for search of house interior via infrared technology), and Riley (requiring warrant for search of cell phone contents).[92] *Carpenter*'s carve outs 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 met with closed courthouse doors. 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 rational basis review, the government's interest in national security arguments often (but not always) trumps. Recent case law has virtually ruled out *Bivens* as a route for plaintiffs alleging abuses in national security cases.

This section will first examine the significant barriers to injunctive relief for victims of discriminatory AI bias, and in so doing, provide arguments for why injunctive relief may be plausible under current case law, if the courts were to reach the merits of claims. It will then briefly address the non-viability of Bivens claims and the need for Congress to create a statutory damages remedy.

1. Barriers to Injunctive Relief

Injunctive relief for equal protection violations and First Amendment discrimination is essential but neither likely nor hardly a fix-all to biased AI or any other national security profiling. There are many barriers to injunctive relief. Some 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*, 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*

[91] See, e.g., In re Directives ?; c.f., the Keith case (showing deference with respect to foreign relations and international security but less deference with respect to purely domestic national security investigations).
[92] See Fourth Amendment discussion in AI and Judges FJC for further discussion.

Fair Admissions, Inc. v. President and Fellows of Harvard College[93]*;* and (5) injunctive relief may be impracticable or impossible for victims of profiling to seek at the time of injury.

(1) 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 treatment is actionable under equal protection law precedent[94]: "To state an equal-protection claim, Plaintiffs must allege (and ultimately prove) "intentional discrimination."[95] 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."[96] 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."[97]

The standard for pleading such "purposeful discrimination" is formidable, requiring "more than 'intent as volition or intent as awareness of consequences.'"[98] 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 . . . . It follows that, to state a claim based on a violation of a clearly established right, respondent must plead sufficient factual matter to show that petitioners 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.[99]

Iqbal alleged that on the basis of his race, religion, and/or national origin, FBI designated him a person of "high interest" and arrested and detained him as part of its investigation of 9/11, that defendants John Ashcroft and Robert Meuller 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

[93] *Students for Fair Admissions, Inc. v. President and Fellows of Harvard College*, 600 U.S. 181 (2023)

[94] I will add in text or footnote the criticisms of the current doctrine, including those of Charles Lawrence, supra note ---, at 322 ("…[R]equiring proof of conscious or intentional motivate 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.")

[95] *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).

[96] *Washington v. Davis*, 426 U.S. 229, 239 (1976).

[97] *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)).

[98] *Id.*

[99] *Id.* at 676-77 (internal citations omitted). [The *Iqbal* case famously followed *Twombly* (200-) in raising the pleading standard – the level of factual matter deemed to be "sufficient" as described above -- that plaintiffs must show to survive a motion to dismiss.]

confinement solely on the basis of "his religion, race, and/or national origin."[100] Ashcroft was alleged to be the "principal architect" of the policy, and Mueller, "instrumental in [its] adoption, promulgation, and implementation."[101]

The Court did not find disparate treatment had been sufficiently pleaded. The Court opined that the September 11 attacks:

> . . . . were perpetrated by 19 Arab Muslim hijackers who counted themselves members in good standing with al Queda, an Islamic fundamentalist group …headed by another Arab Muslim . . . and composed in large part of his Arab Muslim disciples. It should come as no surprise 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.[102]

But of course it is problematic that a search for Islamic *fundamentalists* in al Queda resulted in the detention of "Arab Muslims." The leap from detaining Al Queda fundamentalists to detaining people on the basis of Arab origin and membership in world's second largest religion describes something far different from "incidental" impact or causation. The opinion makes clear that a "disparate, incidental" impact will not be enough to show purposeful discrimination, 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 Timothy McVeigh's religious views 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. Indeed, most scholarship on AI and equal protection accepts this as a given – that biased AI suits will be characterized as disparate impact suits and non-starters.[103] 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 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 and courts accepting the reprehensible argument that because an algorithm is trained on a limited set of input data – the 19 highjackers for 9/11, for extreme example – we should not be surprised that it produces outcomes that overlap on some features, such as "Arab" or "Muslim" or "male," even though those are not the causal or determinative feature – al Queda membership – in predicting terrorist activity. The problem is that the results will be both under- and overinclusive. 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,[104] who

[100] *Id.*

[101] *Id.*

[102] *Id.* at 682 (emphasis added)

[103] Cite examples here. At least one scholar, Stephanie Bornstein, however, helpfully argues in the employment context that biased AI could be challenged under an antistereotyping theory of disparate treatment. Bornstein, Antidiscriminatory Algorithms, 70 ALA. L. REV. 519 (2018).

[104] See, e.g., Sinnar, supra note ___, at 1388-92.

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.

(a) Why using biased AI constitutes disparate treatment

There is room for hope – and good faith 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 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. 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.[105] 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 "over 80% of [NYPD's] 4.4 million stops [between January 2004 and 2012] were of blacks and Hispanics."[106]

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 members of any other religion. . . . . Rather, Plaintiffs' religious affiliation must have been a substantial factor in that different treatment."[107] 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"[108]

[105] *Hassan v. City of New York*, 804 F.3d 277, 285 (3d. Cir. 2015).
[106] *Floyd v. City of New York*, 959 F.Supp.2d 540, 540 (S.D.N.Y. 2013).
[107] *Hassan*, 804 F.3d at 294–95.
[108] *Hassan v. City of New York*, 804 F.3d 277, 294–95 (3d. Cir. 2015); see also *Floyd v. City of N.Y.*, 959 F. Supp. 2d 540, 660-61 (S.D.N.Y. 2013)("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

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.[109] Notably, the court stated "That we might conjure up some non-discriminatory motive to explain the City's alleged conduct is not a valid basis for dismissal."

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 (e.g., at airports, with facial recognition cameras, etc.); 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, etc. 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,*[110] 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,"[111] or if so, that it was a United States rather than foreign operated list.[112]

If it can be shown that an AI model is discriminatory, either through circumstantial evidence as in Hassan, or with access to the actual algorithm, 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,"[113] if not explicitly in its programming, then somewhere

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).")

[109] *Hassan*, 804 F.3d at 295.

[110] C*lapper v. Amnesty International USA*, 568 U.S. 398 (2013)

[111] *Zaidan v. Trump,* 317 F. Supp. 3d 8 (D.D.C. 2018)

[112] *Id.* at __.

[113] *See* Geoffrey R. Stone, Louis M. Seidman, Cass R. Sunstein, Mark V. Tushnet, Pamela S. Karlan, Aziz Z. Huq, and Leah M. Litman, CONSTITUTIONAL LAW, Ninth Ed., 483, ___ (2023)(discussing facially discriminatory laws).

within its workings, such that its outputs fall within those lines. The AI is using 'race' or another category explicitly as the basis for a burden or disadvantage (or advantage). The AI is explicitly fencing people into different groups – it may be explicitly using a language of numbers, not words, but it is still openly fencing them, and the government is adopting that segregating system.

Another advantage, for plaintiffs at least, is that if the AI is facially discriminatory, they need not show discriminatory purpose for courts to apply strict scrutiny.[114] If a government agency uses an AI that when tested produces objectively measurable discriminatory results, 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, etc.

A final advantage of the facially-discriminatory-AI theory is that it fits well into the Court's "autoclassification" theory of equal protection law. The AI is classifying, and that is a constitutional equal protection violation (unless a compelling governmental reason can be shown and the government policy is narrowly tailored, etc.). Moreover, such a theory allows for data engineering or modeling to *reduce* bias. Any engineering or modeling done to reduce bias would be done to 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 can still argue that the government's choice to use the AI shows purposeful discrimination. The government is on notice, as evidenced by its 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.[115]

In *Floyd*, the court held that plaintiffs had shown an intentionally discriminatory application of a facially neutral policy,[116] the second theory outline outlined above for showing disparate treatment. Plaintiffs' "statistical evidence of racial disparities in stops [was] sufficient to show a discriminatory effect."[117] 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 than white people *within* precincts; and they were more likely to use force against Black and Hispanic people than white

[114] *See* Chemerinski, *supra* note ___, at ___.

[115] 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."

[116] *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."

[117] *Id.* at 661

people.[118] They also showed that NYPD officers stopped Black and Hispanic people with less justification than white people.[119] This "statistical evidence of a racially disproportionate impact" was supplemented with significant anecdotal evidence.[120]

AI civil rights litigants might take note. The fine line between non-actionable "disparate impact" and actionable "statistical evidence of racial disparities. . . sufficient to show a discriminatory effect"[121] 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 had discriminatory effect, as well as by anecdotal evidence suggesting intent, such as differences in how officers reported stops. 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; likewise, who was detained in maximum security prisons post 9/11.

Perhaps most helpful for AI plaintiffs, the *Hassan* court empathized 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.[122] 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?'"[123] "Invidious" motive, the 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*."*[124] 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, it concluded, "Thus, 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."[125] 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 'ill will not enmity.'"[126]

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, performance results, etc., are all knowable (and discoverable, if courts so determine). If AI is programmed with suspect categories or proxies, that looks quite intentional. The government is on notice, as evidenced by its own policies, that AI often discriminates. The government has a

[118] *Id.*
[119] *Id.*
[120] *Id.* at 661-62.
[121] *Id.* at 661.
[122] *Hassan*, ____ at 297.
[123] *Id.* at 297 (quoting 1 John William Slamon, Jurisprudence §134 at 398 (7th ed. 1924)
[124] Id. at 297
[125] Hassan at 298
[126] Floyd at 662

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 it knows and uses the AI anyway, that is or at least strongly suggests intentional disparate treatment.[127]

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

Even if a plaintiff successfully pleads facial discrimination or disparate treatment[128] 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."[129]

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 the courts have a history of often, though not always, deferring to the executive

[127] (Counterargument, in DRAFT outline: Ashcroft v. Iqbal, Korematsu, Trump v. Hawaii: both of those cases seem to argue that the govt classification was for another reason, not intent to harm. That argument, however, blends into the government's governmental interest in any scrutiny-based review, not necessarily what plaintiff needs to plead to trigger strict scrutiny. Courts should review the reason later, under strict scrutiny. Still courts might apply the rationale that the government classified for non-discriminatory reasons against plaintiffs sooner in determining what level of review to apply. This is tricky: Iqbal states purposeful discrimination = "not for a neutral, investigative reason but for the purpose of discriminating on account of race, religion, or national origin." Hassan's statement that it need not be malicious, just present, may be an outlier. (Iqbal doesn't, however, say purpose must be malicious, just purposeful discrimination on the basis of race, etc.) Regardless, if I were a government attorney advising on AI, I would not want to be basing my arguments on Korematsu or Trump v. Hawaii – where the evidence of discrimination was clear – nor Ashcroft v. Iqbal, where even the Supreme Court found the allegations were heinous if true.). Another response, more broadly might be found in Margaret Hu's essay, *Digital Interment*, 98 Texas Law Review Online 174-183 (2020). Hu argues we need a legal test that gets to the right question: "Because of the national security framing of the issue, the Court in Trump v. Hawaii focused on the wrong question. It asked whether the President can take necessary action to defend the nation by preventing foreigner entry into the United States. The Court in Korematsu also answered the wrong question. It asked whether the military could take necessary action to defend the nation by preventing sabotage through the containment of potential enemy sympathizers and spies through mass evacuation. In Korematsu, the real question was whether the animating force behind this drastic measure was truly one of national security. Like internment, the real question in Trump v. Hawaii was whether the executive action was based upon prejudice and xenophobia, economic populism and protectionism, and white nationalism." I may do more with this, here or elsewhere.)

[128] Geoffrey R. Stone, Louis M. Seidman, Cass R. Sunstein, Mark V. Tushnet, Pamela S. Karlan, Aziz Z. Huq, and Leah M. Litman, CONSTITUTIONAL LAW, Ninth Ed., 483 (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 if 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."). C.f., Erwin Chemerinsky, CONNOTATIONAL LAW, PRINCIPLES AND POLICIES, Sixth Edition, 776 (2019)("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 governmetn's ability to show a compelling purposes for it.")

[129] *Students for Fair Admissions, Inc. v. President and Fellows of Harvard College*, 600 U.S. 181, 206-07 (2023)

branch in matters of national security.[130] 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*.

On the other hand, Roberts recently wrote that at least with respect to any "race-based" equal protection claims, outside the circumstances of affirmative action (and the repudiated *Koramatsu*), 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." [131] 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.

(3) 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.[132]) 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 either a *Mandel* or rational basis review in the future: 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,"[133] or (2) all "admission and immigration cases that overlap with the area of national security" where Mandel's narrow standard of review "has particular force;"[134] or (3)(perhaps synonymous with 2) 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," [135] or even more broadly. In a footnote, Chief Justice Roberts suggests he might apply "a more constrained standard of review" even to "immigration policies, diplomatic sanctions, and military actions," or even simply

[130] *See, e.g., Trump v. Hawaii (quoting Humanitarian Law Project,* 561 U.S., at 34, 130 S.Ct. 2705 "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.'"); Robert M. Chesney, *National Security Fact Deference*, 95 Va. L. Rev. 1361 (2013).

[131] *Students for Fair Admissions, Inc.*, 600 U.S. at 207 and n.3

[132] *Trump v. Hawaii*, 138 S.Ct. at 2420.

[133] *Trump v. Hawaii*, 138 S.Ct. at 2419.

[134] *Id.*

[135] *Id.* at 2418 (". . . plaintiffs 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 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.")

"the national security and foreign affairs context," though perhaps only if the entry of foreign nationals is involved:

> 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.[136]

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 "the 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," presumably, executive decisions in immigration. 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 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*, Roberts 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 with the executive authority…."

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 set back 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 they 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

[136] *Id.* at n.5 (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." *Id.* at 2439-40 (emphasis added). That is, the Court held only that, in *that* context, a facially neutral policy survives judicial scrutiny if "it can reasonably be understood to result from a justification independent of unconstitutional grounds." *Id.* at 2420. Nothing in the opinion indicates that this "circumscribed inquiry" applies outside of the "national security and foreign affairs context." *Id.* at 2420 n.5.

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?[137]

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

The implications for AI of the 2023 affirmative action case, *Students for Fair Admissions (SFFA)[138],* and the Court's "anticlassification (formal equality)"[139] theory of equal protection law are unclear. 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. As discussed below, it is unclear whether the current Supreme Court, which recently rejected affirmative action under an anticlassification theory of equal protection, will do about this technical problem of AI.

The issue arises when AI programmers see or foresee bias in their models and move to correct it. The Court may not be sympathetic to bias-reduction work if that work is viewed more like "affirmative action" for AI.[140] Rather, it might require AI models to be run on "neutral" data. (That neutrality, of course, would be a myth, especially if historic and current systematic injustices present in the training, testing, and validation data mathematically militated biased results.)

In equal protection challenges to discriminatory AI, this might show up as a defense.
The government might argue that had it removed the bias by "affirmative action" AI, it would have violated the autoclassification 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.

Perhaps *SFFA* can be read the opposite way, against the original, biased AI. If the biased AI bins 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 that is facial classification subject to strict scrutiny. The original, biased AI is the facially discriminatory policy (as argued above). Then, any data engineering or modeling done to reduce bias would be corrective, like desegregation.[141] Reducing bias would look less like affirmative action, which the Court has rejected, and more like an anti-classification 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 AI, rather than the initial design and programming phases. In examining the AI, rather than looking only at the inputs

[137]

[138] *Students for Fair Admissions, Inc. v. President and Fellows of Harvard College*, 600 U.S. 181 (2023)

[139] Jason R. Bent, *Is Algorithmic Affirmative Action Legal?*, 108 GEORGETOWN L. J. 803, 8__ (2020)

[140] Associate Dean Bent, *supra* note 114, wrestles with this problem in the employment law context, ultimately finding ways to justify progressively "race-aware" algorithms.

[141] *See Students for Fair Admissions, Inc.*, 600 U.S. at 207.

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.[142] Rather than looking at the initial "neutral" programming, which is never, in fact neutral, it asks the court to look at the moment the government needs to determine whether to use 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.

(5) 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 84 "aliens" held in the Metropolitan Detention Center (MDC) in Brooklyn while the FBI investigated them post 9-11; as in *Ashcroft v. Iqbal*, the plaintiffs 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. [143] 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.[144] 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. As Justice Kennedy summarized:

> The complaint includes these allegations: Conditions in the Unit were harsh. Pursuant to official Bureau of Prisons policy, detainees were held in " 'tiny cells for over 23 hours a day.' " . . . Lights in the cells were left on 24 hours. Detainees had little opportunity for exercise or recreation. They were forbidden to keep anything in their cells, even basic hygiene products such as soap or a tooth-brush. When removed from the cells for any reason, they were shackled and escorted by four guards. They were denied access to most forms of communication with the outside world. And they were strip searched often—anytime they were moved, as well as at random in their cells.

As Justice Breyer wrote in dissent,

> Some of the plaintiffs allege that for two or three months they were subject to a "communications blackout"; that the prison "staff did not permit them visitors, legal or social telephone calls, or mail"; that their families and attorneys did not know where they were being held; that they could not receive visits from their

[142] 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.

[143] [[[Verify: " involves…"individuals who were caught up in the post–9/11 investigation even though they were unquestionably never involved in terrorist activity. " *Turkmen v. Hasty*, 789 F.3d 218, 224 (2d Cir. 2015). All 84 or just those 8?]]]

[144]

> attorneys; that subsequently their lawyers could call them only once a week; and that some or all of the defendants "interfered with the detainees' effective access to legal counsel."[145]

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."

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 injunctive relief may not be available in the actionable moment, nor will it address past harms. After-the-fact monetary damages must also be available.

C. *Bivens* Presents No Viable Option for Post-Harm Relief for Discriminatory AI.

Despite the horrendous nature of the allegations in *Ziglar v. Abbassi*, 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." [146] 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."[147] The 2022 Supreme Court decision *Egbert v. Boule* effectively foreclosed national security Bivens cases.[148]

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.[149] 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, etc.? I would argue that at a minimum any contractor who provides and profits from a discriminatory model should be accountable. Like the government, the contractor has an obligation to test and

[145]

[146] Dycus, Banks, Raven-Hansen, and Vladeck, supra note ___, at 150.

[147] 137 S. Ct. at 1861; see Dycus, Banks, Raven-Hansen, and Vladeck, supra note ___ at 150-1.

[148] Cite to Dycus, Banks, etc., supplement or teachers manual?

[149] *See id*. at 150-51 ("One important consequence of *Ziglar* is that the denial of a cause of action for damages is fatal to a claim for any *retrospective* judicial relief whatever, even for the most egregious and clearly established constitutional violations. But won't such a holding confer a form of "absolute" immunity on federal government officers who never have to fear damages liability for constitutional violations in national security litigation (if not more generally)?"

monitor its AI at least up until the point of sale (and possibly afterward, if it is administering the government program or if it finds bias in its models employed in other contexts). Government officials who request or approve of the use of discriminatory AI should be liable. Decisions to use AI should be made at the highest levels because of their grave consequences for humanity.

[Problematically, *Iqbal* rejected the idea of supervisory liability in the Bivens context. I will add discussion here.]

(I would argue that at a minimum any contractor who provides and profits from a discriminatory model should 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 it is administering the government program or if it finds bias in its models employed in other contexts).

In light of the Court's stated deference to the political branches, Congress should consider creating a cause of action for damages for AI discriminatory harms. [Does Congress also have the authority to dictate that high level officials are accountable, despite the Court's rejection of supervisory liability in Bivens cases? A congressionally create cause of action would not be a Bivens claim, technically.]

D. Watchlisting: AI Risks for Fifth Amendment Due Process

[I have written about watchlisting and AI elsewhere and may incorporate some of that analysis here.]

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[150] – 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. This will make it especially challenging for plaintiffs to show injury for standing, for example, as in the non-AI cases Clapper v. Amnesty International,[151] where plaintiffs could not establish standing because the Court thought their injuries under the classified 702 surveillance program too speculative, or Halkin v. Helms,[152] where the government asserted states secrets, blocking discovery necessary for plaintiffs to assert standing.

[[Add: Iqbal Twombly standard;

Qualified Immunity (easy to avoid both legality question and liability question when no case law on AI and arguable nature of questions on the merits; another reason for Congress to legislate and answer questions by providing causes of action for damages relief for disparate treatment and injunctive and damages relief for disparate impact)

[150] See generally, Dycus, Banks, Raven-Hansen, Vladeck, supra note __, ch. 5.

[151] Clapper v. Amnesty Int'l USA, 568 U.S. 398 (2013)

[152] Halkin v. Helms, 690 F.2d 977 (C.A.D.C. 1982)

States secrets [153]/ classified v. due process arguments why AI must be transparently used Fact deference even if the get to the merits, as in Trump v. Hawaii, examples in Chesney article]]

III. Executive Policies Relevant to Discriminatory AI

[[Outline/DRAFT: Some good, some bad, all reversible by the next administration. And all done behind classified doors, where accountability is only as good as the ethics and fortitude of the individuals in the room, and were unconscious biases, including limited experiential perspective where the room insufficiently diverse, is even harder to unmask, even for good faith actors.]]

A. Department of Justice Guidance for Federal Law Enforcement Agencies Regarding the Use of Race, Ethnicity, Gender, National Origin, Religion, Sexual Orientation, Gender Identify, and Disability[154]

In May 2023 the Department of Justice updated its Guidance for federal law enforcement agencies on the use of suspect classifications. As Professor Faizel Patel writes, the Guidance remains too permissive of profiling: "although the new rules include some improvements (e.g., explicitly covering intelligence and other activities supporting law enforcement), they continue 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."[155]

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 Margaret Hu has argued, between law enforcement and national security data regimes.

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

[153] Discuss Bobby Chesney on doctrinal confusion and Shirin Sinnar on masked harms from recent Harvard Law Review Forum

[154] U.S. Department of Justice, "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/2023-06/Guidance%20for%20Federal%20LEAs%20on%20the%20Use%20of%20Protected%20Characteristics_FINAL%205.25.23_508.pdf

[155] 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 and 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)

> 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."[156]

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 the "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 group" using ethnicity and national origin, the results would look like the propensity profiling 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.

B. Intelligence Community Principles and Ethical Framework for AI.

The Intelligence Ethics Framework offers excellent questions for technologists and attorneys to ask. It is very detailed for such a document. 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, including components used in AI development? 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?

[156] U.S. Department of Justice, Guidance, supra note 128, at ___.

- 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?[157]

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.[158] Likewise to 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 also unrealistic to expect that they will perfectly understand and account for the potential to particular groups.[159] For an agency attorney, for example who already has an endlessly busy job of dealing with day-to-day agency operations, administration, litigation, etc., to also be the gatekeeper of the use and maintenance of complex AI systems, is unrealistic. But right now, those attorneys, the officials they advise, and perhaps a few civil rights and civil liberties officers, are all we've got.

C. New Executive Order on AI:

In October 2023, the Biden Administration released Executive Order (E.O.) 14110 on Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence.[160]

Section 10.1(b)(iv) (for government use of AI other than in national security systems): Requires, for Government uses of AI that impact people's rights or safety, that agencies follow, "where appropriate," these practices "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,"

[[Outline/DRAFT:

good to "require" unclassified AI systems be disclosed publicly but nothing on classified systems; simply a requirement for a national security memo on AI use (see section II, above);

[157] https://www.intelligence.gov/artificial-intelligence-ethics-framework-for-the-intelligence-community#Purpose

[158] 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 at sentencing, where no "prediction" should be made, but courts have nonetheless employed them that way (see *Wisconsin v. Loomis*); there are of course many, 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.

[159] Cite to CRT articles; if not already discussed above, do so here

[160] E.0. 14110 on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (Oct. 30, 2023), https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/

--Transparency is needed for even classified AI uses – hard to figure out what those uses are event to write this paper]]

also told to put generative AI into use, but without any guidance about how to do that consistently with civil rights and civil liberties (cite lawyer who pointed that out to me with permission)]]

--Faiza Patel and Patrick Toomey argue bad to have two systems, one for other agencies and one for national security agencies

--Margaret Hu's point that various systems will be increasingly linked by AI is another reason why it's bad to have two systems

--Civil Rights Civil Liberties Officer section?

IV. Recommendations to Mitigate and Challenge Discriminatory AI

A. 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[161] and Egbert[162] 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 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. If Congress took that last approach, it might caveat that by providing for a specific damages remedy with respect to algorithmic discrimination, it did not intend to preclude the courts from remedying constitutional torts under their equitable powers in other contexts.

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

[161] *Ziglar v. Abbasi*, 137 S.Ct. 1843, 1857-58, 1863 (U.S., 2017)("If *Bivens* liability were to be imposed, high officers who face personal liability for damages might refrain from taking urgent and lawful action in a time of crisis. And, as already noted, the costs and difficulties of later litigation might intrude upon and interfere with the proper exercise of their office. On the other side of the balance, the very fact that some executive actions have the sweeping potential to affect the liberty of so many is a reason to consider proper means to impose restraint and to provide some redress from injury. There is therefore a balance to be struck, in situations like this one, between deterring constitutional violations and freeing high officials to make the lawful decisions necessary to protect the Nation in times of great peril. . . . The proper balance is one for the Congress, not the Judiciary, to undertake.")

[162]

damages relief in disparate impact suits against federal, state, and local government actors using discriminatory AI.[163] The Court has suggested that Congress has the authority to do so in *Washington v. Davis*:

> A rule that a statute designed to serve neutral ends is nevertheless invalid, absent compelling justification, if in practice it benefits or burdens one race more than another would be far reaching and would raise serious questions about, and perhaps invalidate, a whole range of tax, welfare, public service, regulatory, and licensing statutes that may be more burdensome to the poor and to the average black than to the more affluent white.
>
> Given that rule, such consequences would perhaps be likely to follow. However, in our view, extension of the rule beyond those areas where it is already applicable by reason of statute, such as in the field of public employment, should await legislative prescription.

As 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."[164] 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.[165]

In providing either monetary damages for disparate treatment by AI (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 additional oversight mechanisms, especially a court, similar to the FISA Court, for AI-related civil rights and liberties.

Draft Outline:

a. Ashley Deeks has an interesting proposal worth exploring that there should be a covert action like congressional approval system for AI – her focus seems to be largely on safety and international repercussions, traditional covert-action type concerns.
b. That or an equivalent congressional approval system for AI might also include briefing on disparate impact, disparate treatment, and other related harms
c. Civil rights and civil liberties concerns would, in theory, be best addressed by a court. Make a FISA court equivalent for constitutional rights concerns about government AI systems. That court would look specifically for equal protection, first amendment, due process concerns for each classified AI model.

[163] Congress might also consider authorizing disparate impact suits against private actors, though the Court might be less likely to uphold it.
[164] Chemerinsky, supra note __, at 770.
[165] *Id.*

Agencies would be required to file for approval before implementing an AI system or model, and also to check in at regular intervals during its deployment to provide monitoring data. Might consider making this an executive court like the new data review court if that would promote the court getting to the merits.

d. Expand mandate of PCLOB or another executive agency to also act as oversight (perhaps already in PCLOB's mandate – but call upon PCLOB or another agency to look specifically at AI bias)
e. Require agency IGs and civil rights/liberties officers to conduct oversight; require those officers, attorneys, and managers to report to IOB or a newly—created similar entity any suspected disparate treatments or disparate impacts by AI. Reporting should be quarterly or more often.

B. For Litigants (and Courts)

Draft outline: See Strategies outlined above, but in sum: argue that AI is methodical and objectively measurable (though not objective). Therefore, regular Fourth Amendment concerns about reasonableness, and not *Whren*, apply. Under equal protection doctrine, any government use of biased AI is a facially discriminatory policy, or at the very least constitutes purposeful disparate treatment. 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, to the court and cleared counsel, at a minimum. (If the AI is not transparent, that is on the developer / govt – get better technology or pay the price in a lawsuit.). It must be as unbiased as possible – and if bias cannot be mitigated sufficiently, then it should not be used. [How to determine what level of mitigation is sufficient the ultimate hard question – perhaps none is.]

C. For Executive Attorneys

Draft outline: As ever, government attorneys must be guardians of constitutional rights where the courts and Congress are not yet. Government attorneys should learn AI fundamentals and talk with technologists throughout the lifecyle of the AI, from inception to end of use, asking pointed questions at every step along the way to test for and weed out bias. Consider 'affirmative action' or what Stephanie Bornstein smartly calls "antidiscriminatory AI" where possible. Set up procedures for regular monitoring for bias. And thank you, for doing this hard and thankless and detailed task.

Conclusion