

Overcoming Racial Harms to Democracy from Artificial Intelligence

*Spencer Overton**

ABSTRACT: While the United States is becoming more racially diverse, generative artificial intelligence and related technologies threaten to undermine truly representative democracy. Left unchecked, AI will exacerbate already substantial existing challenges, such as racial polarization, cultural anxiety, antidemocratic attitudes, racial vote dilution, and voter suppression. Synthetic video and audio (“deepfakes”) receive the bulk of popular attention—but are just the tip of the iceberg. Microtargeting of racially tailored disinformation, racial bias in automated election administration, racially targeted cyberattacks, and AI-powered surveillance that chills racial justice claims are just a few examples of how AI is threatening democracy. Unfortunately, existing laws—including the Voting Rights Act—are unlikely to address the challenges. These problems, however, are not insurmountable if policymakers, activists,

* Patricia Roberts Harris Research Professor of Law and Faculty Director of the GW Law School’s Multiracial Democracy Project, George Washington University. Early exchanges with Kelly Born, Rita Bosworth, David Brody, Danielle Brown, Danielle Keats Citron, Julia Rhodes Davis, Torey Dolan, Jessica Eaglin, Eric Easter, Farbod Faraji, David Fontana, Chanelle Hardy, Woodrow Hartzog, Matthew Hindman, Orin Kerr, Pauline Kim, Eric Loeb, Stefan McCabe, Alan Morrison, Dawn Nunziato, Ngozi Okidegbe, Nate Persily, Dick Pierce, Ethan Porter, Ann Ravel, Blake Reid, Barak Richman, Ethan Roeder, Marc Rotenberg, Nicole Schneidman, Bruce Schneier, Jon Siegel, Alicia Solow-Niederman, Olivier Sylvain, Rebekah Tromble, Charles Tyler, Tania Valdez, Paul Waters, Kate Weisburd, and Glen Weyl helped develop the ideas in this Article. The Article has also benefited from my presentations on AI and democracy at the “Advances in Deepfake Technology” hearing of U.S. House Committee on Oversight and Accountability Committee’s Cybersecurity, Information Technology, and Government Innovation Subcommittee; the Association of American Law Schools 2024 Annual Meeting; the Carnegie Endowment for International Peace’s Workshop on AI and Democracy; GW Law School’s Faculty Workshop Series; the Harvard Kennedy School Ash Center’s Interdisciplinary Workshop on Reimagining Democracy; the Harvard Law School Chapter of the American Constitution Society; the Knight Foundation’s INFORMED Conference; the Stanford Cyber Policy Center; and the Washington and Lee Law School’s Lara D. Gass Symposium. Rick Banks, Robert Brauneis, Jonathan Cedarbaum, Tom Colby, Josh Douglas, Moon Duchin, Mary Anne Franks, Yosef Getachew, Ruth Greenwood, Rick Hasen, Aziz Huq, Meetal Jain, Kevin Johnson, Justin Levitt, Laura MacCleery, Dayna Matthew, Alondra Nelson, Ali Noorani, Larry Norden, Mekela Panditharatne, Robert Post, Nick Stephanopoulos, Ed Swaine, Tova Wang, Dave Willner, and Fane Wolfer read earlier drafts of this Article and provided helpful comments. Priyanka Mara, Gideon Epstein, and Eleonora Viotto provided invaluable research assistance. I am particularly grateful to the Democracy Fund, Open Society Foundations, the David and Lucile Packard Foundation, and the Rockefeller Brothers Fund for their generous support of the GW Multiracial Democracy Project, which has allowed me to focus on this research.

and technology companies act now. This Article asserts that AI should be regulated to facilitate a racially inclusive democracy, proposes novel principles that provide a framework to regulate AI, and offers specific policy interventions to illustrate the implementation of the principles. Even though race is the most significant demographic factor that shapes voting patterns in the United States, this is the first article to comprehensively identify the racial harms to democracy posed by AI and offer a way forward.

- INTRODUCTION 807
- I. AI AND LAW SHOULD FACILITATE RACIALLY INCLUSIVE DEMOCRACY 810
 - A. *MULTIRACIAL DEMOCRACY IN AMERICA*..... 810
 - B. *INCREASING RACIAL POLARIZATION AND ANTIDEMOCRATIC ATTITUDES* 814
- II. AN OVERVIEW OF AI AND DEMOCRACY 816
- III. AI’S RACIAL HARMS TO DEMOCRACY 821
 - A. *RACE AND INFORMATION INTEGRITY* 823
 - 1. Racial Impersonation and Infiltration..... 823
 - 2. Microtargeting, Manipulative Chatbots, and Cultural Conquest..... 829
 - 3. Fueling Cultural Anxiety 833
 - B. *AI MODEL DESIGN CAN ENTRENCH RACIAL HIERARCHY*..... 835
 - 1. AI Models Can Minimize the Perspectives of People of Color 837
 - 2. Language Models Can Suppress Language Minority Participation..... 839
 - 3. Racially Biased Content Moderation 843
 - C. *THE CHILLING EFFECTS OF AI SURVEILLANCE IN OVERPOLICED AND IMMIGRANT COMMUNITIES*..... 844
 - D. *RACIAL ENTRENCHMENT THROUGH ELECTION STRUCTURES*..... 848
 - 1. Automating Bias in Election Administration 848
 - 2. Racially Targeted Attacks on Local Election Offices..... 851
- IV. PRINCIPLES TO OVERCOME AI HARMS TO DEMOCRACY..... 854
 - A. *ANTICIPATE RACIAL HARMS TO DEMOCRACY* 859
 - B. *FACILITATE PLURALISM AND PREVENT ALGORITHMIC DISCRIMINATION*..... 862
 - C. *MITIGATE RACIAL DISINFORMATION AND MANIPULATION* 863
 - D. *PROVIDE MEANINGFUL ACCOUNTABILITY*..... 865
- CONCLUSION 866

INTRODUCTION

By the year 2050 there will be no majority ethnic group in the United States. Although the nation has made great progress toward becoming a racially inclusive democracy, key obstacles remain. As the nation has become more diverse, racial polarization and cultural anxiety have increased. Unfortunately, another major but to date underappreciated threat has emerged at the intersection of race and democracy: artificial intelligence (“AI”). This Article addresses that threat.

While synthetic video and audio (“deepfakes”) receive the bulk of popular attention, generative AI and related technologies are transforming all aspects of our electoral system and pose significant challenges to the future of a racially inclusive democracy. AI could empower politicians to iterate billions of datapoints to find hidden patterns and develop customized messages—including those seemingly unrelated to voting—designed to covertly manipulate the cultural and political identity of individual voters of color or reduce voter turnout in communities of color.¹ AI will empower hackers to deploy cyberattacks and swarms of frivolous open-record requests targeted to incapacitate local election offices that serve large populations of voters of color.² Local law enforcement’s warrantless deployment of AI to surveil social media and mobile phone data will continue to chill demonstrations for police reform by Black Lives Matter activists.³

Even absent intentional discrimination, bias and other flaws embedded in AI datasets used to create content, moderate content, detect deepfakes, maintain voter rolls, verify mail-in ballot signatures, provide adequate language assistance in voting, and perform a host of other tasks could replicate disadvantages and fortify racial and cultural hierarchy in elections and policymaking well into the future.⁴

The challenges arise both from the bias of AI datasets and the limitations in the design of AI to adequately serve all in a pluralistic, diverse society. Because AI models typically address diversity in mathematical ways by defaulting to averages or dominant patterns, they often fail to recognize different perspectives and needs—including those of communities of color in the context of voting and democracy. The current design of many AI models effectively automates homogenization and reifies racial stereotypes, which is inconsistent with respect for and the coexistence of both factual information and diverse interests and viewpoints in a liberal democracy. Technologists have only recently started to acknowledge these design flaws.⁵

1. *See infra* Section III.A.2.

2. *See infra* Section III.D.2.

3. *See infra* Section III.C.

4. *See infra* Section III.B.

5. *See generally* Taylor Sorensen et al., *A Roadmap to Pluralistic Alignment*, 235 *PROC. MACH. LEARNING* 46280 (2024) (discussing how designing AI systems to serve people with diverse values and perspectives remains an open research question and proposing a roadmap to pluralistic alignment of these systems); Mitchell L. Gordon et al., *Jury Learning: Integrating Dissenting Voices into Machine Learning Models*, 2022 *PROC. CHI CONF. ON HUM. FACTORS COMPUTING SYS.* 115

The harms of AI to deliberation, participation, and representation are often more difficult to detect and regulate when they disproportionately affect groups with less social, political, and economic influence—including many communities of color. The homogeneity of those who develop the tools and govern tech companies and the failure to prioritize the unique ways in which many communities of color experience AI technologies only compound the antidemocratic nature of the harms.

This Article does not take the position that AI inevitably spells the end of democracy. Rather than ignore AI's racial and cultural implications for democracy, we must anticipate and mitigate the threats it poses. A central goal of AI and the law could be—and should be—to facilitate our transition to a well-functioning, inclusive, pluralistic democracy—one that respects both identity and individual autonomy and enables cross-group engagement, coalition building, and collective well-being.⁶ Artificial intelligence and future laws that regulate it should encourage incumbent politicians to respond to demographic change through fair representation rather than entrenching power by stoking racial animosity and manipulating racial turnout.

Even though race is the most significant demographic factor that shapes voting patterns in the United States,⁷ this Article is the first to comprehensively examine the racial challenges of AI for democracy. Other scholars have started to analyze the implications of AI for democratic deliberation broadly⁸ and

(introducing a supervised machine learning approach called “jury learning” to resolve disagreement about ground truth labels through the metaphor of a jury); Taylor Sorensen et al., *Value Kaleidoscope: Engaging AI with Pluralistic Human Values, Rights, and Duties*, 38 AAAI CONF. ON AI 19937 (2024) [hereinafter *Value Kaleidoscope*] (discussing the benefit of an explicit, modular, and interpretable approach to value pluralism to improve AI systems to better reflect the values, rights, and duties behind human decision-making).

6. See Joel K. Goldstein, *Justice Brandeis and Civic Duty in a Pluralistic Society*, 33 TOURO L. REV. 105, 106 (2017) (indicating that “[Justice] Brandeis’ conception of duty (and rights) coincided with and was linked to his recognition that America was a pluralistic community, and to his commitment to the ideal that America should be open to, respectful of, and welcoming of, diverse people”).

7. See *infra* Section I.B.

8. See, e.g., Archon Fung, Lawrence Lessig & The Conversation US, *How AI Could Take over Elections—and Undermine Democracy*, SCI. AM. (June 7, 2023), <https://www.scientificamerican.com/article/how-ai-could-take-over-elections-and-undermine-democracy> [<https://perma.cc/DU6A-XQ6Y>] (discussing whether “organizations [could] use artificial intelligence language models such as ChatGPT to induce voters to behave in specific ways”); Karl Manheim & Lyric Kaplan, *Artificial Intelligence: Risks to Privacy and Democracy*, 21 YALE J.L. & TECH. 106, 111 (2019) (exploring “present and predicted dangers that AI poses to core democratic principles of privacy, autonomy, equality, the political process, and the rule of law”); Darrell M. West, *How AI Will Transform the 2024 Elections*, BROOKINGS (May 3, 2023), <https://www.brookings.edu/articles/how-ai-will-transform-the-2024-elections> [<https://perma.cc/5PXX-LZ8S>] (discussing how “advances in digital technology provide new and faster tools for political messaging and could have a profound impact on how voters, politicians, and reporters see the candidates and the campaign”); Jonathan Zittrain, *Engineering an Election*, 127 HARV. L. REV. F. 335, 336 (2014) (discussing “digital gerrymandering: the selective presentation of information by an intermediary to meet its agenda rather than to serve its users”); ELIZABETH SEGER, *GENERATIVE AI AND DEMOCRACY: IMPACTS AND INTERVENTIONS* 4 (2024) (emphasizing “the additional ‘marginal risk’ generative AI poses to democracy”); Andreas Jungheer, *Artificial Intelligence and Democracy: A Conceptual Framework*, SOC. MEDIA + SOC’Y 3 (July 16, 2023), <https://journals.sagepub.com/doi/10.1177/20563051231186353> [<https://perma.cc/JE3T-7SAS>] (providing an interdisciplinary framework for a focused

the dangers to elections of AI-enabled deepfakes.⁹ Recognizing the growing significance of AI, demographic change, cultural anxiety, antidemocratic sentiment, and a U.S. Supreme Court increasingly hostile to traditional voting rights protections, this Article provides an essential first step in developing legal structures and private-sector practices to secure representative democracy for future generations in the United States.

Technology is evolving rapidly, and we may not be able to completely eliminate every challenge. But by acting now, advocates, policymakers, and the private sector can contain the most severe harms from these new tools. This Article introduces a framework to approach AI in the context of our increasingly diverse democracy, catalogs key challenges posed by AI, and crafts principles to mitigate these challenges. Future scholarship will isolate and analyze individual problems and propose detailed legal solutions. Recognizing that AI can also be used to mobilize underrepresented voters, expand the capacity of underfunded community organizations and candidates of color, and facilitate cross-racial coalition building, future scholarship will also explore how law can incentivize the development and use of AI applications to enhance racially inclusive democracy.¹⁰

This Article does not grapple with the currently unknown challenges from future generations of artificial intelligence that could become independent of human control. Their implications for multiracial democracy are speculative at best. At this moment, the priority should be identifying, understanding, and taking action to contain the challenges to multiracial democracy that are most likely to arise from existing and emerging AI technologies.

Part I of this Article describes how the United States has made great strides over the past six decades in becoming a racially inclusive democracy. Still, significant challenges remain, including increasing racial polarization, cultural anxiety, and antidemocratic attitudes. Part II provides a brief overview of artificial intelligence and emerging applications of the technology in the

analysis and discussion of AI's likely impact on democracy). See generally Danielle Allen & E. Glen Weyl, *The Real Dangers of Generative AI*, 35 J. DEMOCRACY 147 (2024) (highlighting the possible biases and “singularity” risks of generative foundational models); Paul Nemitz, *Constitutional Democracy and Technology in the Age of Artificial Intelligence*, 376 PHIL. TRANSACTIONS ROYAL SOC. A 1 (2018) (discussing AI and digital power concentration, and the threats they pose to democracy and functioning markets); Iason Gabriel, *Toward a Theory of Justice for Artificial Intelligence*, 151 DAEDALUS J. AM. ACAD. ARTS & SCIS. 218 (2022) (examining the link between AI and distributive justice); PRATHM JUNEJA, ARTIFICIAL INTELLIGENCE FOR ELECTORAL MANAGEMENT (2024), <https://www.idea.int/sites/default/files/2024-04/artificial-intelligence-for-electoral-management.pdf> [<https://perma.cc/M7XS-DTZ2>] (exploring how Electoral Management Bodies (“EMBs”) and non-EMB actors can use AI to impact elections).

9. See Robert Chesney & Danielle Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 CALIF. L. REV. 1753, 1778 (2019); Rebecca Green, *Counterfeit Campaign Speech*, 70 HASTINGS L.J. 1445, 1447 (2019); Richard L. Hasen, *Deep Fakes, Bots, and Siloed Justices: American Election Law in a “Post-Truth” World*, 64 ST. LOUIS U. L.J. 535, 537–65 (2020); Russell L. Weaver, *Fake News (& Deep Fakes) and Democratic Discourse*, 24 J. TECH. L. & POL’Y 35, 37–41 (2020). An important student note has flagged the challenges of AI in automating election administration. See Sarah M.L. Bender, Note, *Algorithmic Elections*, 121 MICH. L. REV. 489, 489 (2022).

10. See generally Spencer Overton, *Analyzing Benefits of Artificial Intelligence to Racially Inclusive Democracy* (forthcoming 2025) (unpublished manuscript) (on file with author).

context of democracy. Part III maps many of the ways in which AI may be used to cause racial harms to democracy. Recognizing that existing legal protections are unlikely to address these challenges, Part IV proposes four key principles for those who develop, deploy, and regulate AI to tackle the challenges and support an inclusive democracy that reflects the will of our increasingly diverse population.

I. AI AND LAW SHOULD FACILITATE RACIALLY INCLUSIVE DEMOCRACY

Although people of color remain underrepresented, in the past sixty years the United States has made significant progress toward becoming a racially inclusive, pluralistic democracy. Some challenges have emerged, however. For example, racially polarized voting, cultural anxiety, and antidemocratic attitudes are increasing. These challenges provide incentives for some political operatives to entrench their power by using AI-powered applications to manipulate election practices and racial turnout. Even when AI-powered tools inadvertently erect barriers to political participation for communities of color, the tools obstruct our nation's progress toward key constitutional principles of equal representation—one person, one vote, for instance, and government of, by, and for the people.¹¹

Artificial intelligence and the laws that regulate it should facilitate a well-functioning, inclusive multiracial democracy that respects identity and individual autonomy. They should also provide opportunities for groups to evolve, recognize common interests, and form new coalitions to tackle pressing challenges. AI and the law should support a democracy that incentivizes political operatives to respond to demographic change rather than acquiring and entrenching power by stoking racial animosity and manipulating election rules.

A. MULTIRACIAL DEMOCRACY IN AMERICA

“[T]he United States is currently undergoing an extraordinary transition to multiracial democracy.”¹² A multiracial democracy is a liberal democracy with a racially and ethnically diverse population that protects the political rights of individual adults of all groups. As Steven Levitsky wrote:

Modern democracy . . . is a system of regular free and fair elections in which all adults possess the right to vote as well as basic civil liberties like freedom of speech, press, association, and protest. A *multiracial democracy* is simply a democracy in a diverse society in which those basic rights are universally protected across ethnic groups; in other

11. *Reynolds v. Sims*, 377 U.S. 533, 558 (1964) (“The conception of political equality from the Declaration of Independence, to Lincoln’s Gettysburg Address, to the Fifteenth, Seventeenth, and Nineteenth Amendments can mean only one thing—one person, one vote.” (quoting *Gray v. Sanders*, 372 U.S. 368, 381 (1963))).

12. Steven Levitsky, *The Third Founding: The Rise of Multiracial Democracy and the Authoritarian Reaction Against It*, 110 CALIF. L. REV. 1991, 1991 (2022).

words, where the rights of individuals of all ethnic groups are protected equally.¹³

Meaningful participation in liberal democracy serves several important functions,¹⁴ particularly in a multiracial society. Participation allows people to assemble, compare their individual experiences, identify collective challenges, and develop solutions.¹⁵

It also allows them to build coalitions with diverse groups, petition government and private-sector actors about challenges and solutions, and create incentives for those petitioned to respond.¹⁶ By exposing decisionmakers to a diverse array of ideas and perspectives, well-functioning multiracial democracies allow government officials and other powerful actors to make more fully informed choices, thanks to more robust deliberation.¹⁷ Participation furthers autonomy, fulfillment, and self-determination by allowing individuals to collaborate with others to help shape the decisions that affect their lives—thereby reducing isolation and subjugation.¹⁸ A well-functioning multiracial democracy also enhances the legitimacy of government and private-sector decisions and the likelihood of compliance with those decisions.¹⁹

13. *Id.* (emphasis added); see also Allen & Weyl, *supra* note 8, at 147 (defining “plural societies” as “free and democratic societies operating under conditions of social diversity”); Ashley Jardina & Robert Mickey, *White Racial Solidarity and Opposition to American Democracy*, 66*Q* ANNALS AM. ACAD. POL. & SOC. SCI. 79, 81 (2022) (“Democracies feature free and fair elections, universal adult suffrage, guaranteed civil rights and civil liberties, and the rule of law.”).

14. See Spencer Overton, *The Participation Interest*, 100 GEO. L.J. 1259, 1274 (2012) (describing the utilitarian benefits of democratic participation).

15. See *id.*

16. See SIDNEY VERBA, KAY LEHMAN SCHLOZMAN & HENRY E. BRADY, VOICE AND EQUALITY: CIVIC VOLUNTARISM IN AMERICAN POLITICS 1, 110, 112 (1995).

17. See Walter A. Rosenbaum, *Public Involvement as Reform and Ritual: The Development of Federal Participation Programs*, in CITIZEN PARTICIPATION IN AMERICA: ESSAYS ON THE STATE OF THE ART 81, 87 (Stuart Langton ed., 1978).

18. See Frank I. Michelman, *Conceptions of Democracy in American Constitutional Argument: Voting Rights*, 41 FLA. L. REV. 443, 451 (1989); see also C.B. MACPHERSON, THE LIFE AND TIMES OF LIBERAL DEMOCRACY 47–48, 51–52 (1977) (discussing different models of democracy).

19. MARY GRISEZ KWEIT & ROBERT W. KWEIT, IMPLEMENTING CITIZEN PARTICIPATION IN A BUREAUCRATIC SOCIETY: A CONTINGENCY APPROACH 132 (1981) (providing examples of citizen participation and presenting the hypothesis that “[t]he more satisfied the citizens are with participation, the more trusting and efficacious they will be”); Luis Fuentes-Rohwer, *The Emptiness of Majority Rule*, 1 MICH. J. RACE & L. 195, 201 (1996).

While the U.S. Constitution,²⁰ state voting restrictions,²¹ and immigration and citizenship laws²² originally entrenched white rule, over the past six decades the United States has moved toward becoming a multiracial democracy in large part due to two laws Congress passed in 1965.²³ The Voting Rights Act initially outlawed literacy tests, allowed for litigation to challenge discriminatory voting practices, and required that certain jurisdictions preclear changes to their voting practices with federal officials.²⁴ The law was enhanced in 1975 to expand the preclearance process and require bilingual election materials in jurisdictions with large language minority populations (speakers of Spanish, Asian, Native American, and Alaskan Native languages).²⁵ The Immigration and Naturalization Act of 1965 explicitly barred racial discrimination in the immigration admissions process and repealed quotas based on race and

20. The U.S. Constitution subsidized slavery by: (1) inflating the power of slaveholding states in the U.S. House and in presidential elections by counting three-fifths of their enslaved populations for apportionment purposes, see U.S. CONST. art. I, § 2, cl. 3; *id.* art. II, § 1, cl. 3; (2) initially preventing Congress from prohibiting states from importing Black people to serve as slaves until 1808, see *id.* art. I, § 9, cl. 1; and (3) extending rights to enslavers to capture Black people who had escaped to free states, see *id.* art. IV, § 2, cl. 3.

21. Almost all states—including those outside of the South—limited voting to white males or would eventually do so. See Eric Ledell Smith, *The End of Black Voting Rights in Pennsylvania: African Americans and the Pennsylvania Constitutional Convention of 1837-1838*, 65 PA. HIST. 279, 279 (1998).

22. Laws regulating citizenship and immigration gerrymandered the racial composition of the U.S. population in a way that ensured white democratic majorities that persist to this day. See, e.g., Naturalization Act of 1790, ch. 3, 1 Stat. 103, 103 (explicitly limiting U.S. citizenship to “free white persons”); National Origins Act of 1924, Pub. L. No. 68-139, 43 Stat. 153 (establishing immigration quotas that heavily favored immigrants from western and northern Europe, and completely banning immigrants from Asia). Congress also delayed statehood—and the accompanying U.S. House and Senate representation—to the Oklahoma, New Mexico, Arizona, Hawaii, and Alaska territories in part because their residents were insufficiently white. See Stephanie M. Wildman, Jean Stefancic, Richard Delgado, Angela P. Harris & Juan F. Perea, *Race and Races: Cases and Resources for a Diverse America* 293 (Santa Clara Univ. Sch. of L., Working Paper No. 07-25, 2007).

23. The first widespread attempt to make the United States a multiracial democracy in the aftermath of the Civil War was significant, but it was short-lived. Due to the Reconstruction Amendments and federal legislation, newly enfranchised Black southern voters (almost completely Republican) initially elected Black Americans to over two thousand local, state, and federal offices. Federal retreat from the South in the 1870s, however, allowed white southern Democrats to take over southern politics by using violence and voting restrictions to purge Black voters and their representatives from the democratic process. See ALEXANDER KEYSSAR, *THE RIGHT TO VOTE: THE CONTESTED HISTORY OF DEMOCRACY IN THE UNITED STATES* 53-76 (2000); Franita Tolson, “*In Whom Is the Right of Suffrage?*”: *The Reconstruction Acts as Sources of Constitutional Meaning*, 169 U. PA. L. REV. 2041, 2045 (2021); ERIC FONER, *RECONSTRUCTION: AMERICA’S UNFINISHED REVOLUTION, 1863-1877*, at 575-601 (Henry Steele Commager & Richard B. Morris eds., 1988).

24. See Voting Rights Act of 1965, Pub. L. No. 89-110, § 2, 79 Stat. 437 (codified as 52 U.S.C. § 10301(a) (2018)); see also Abhay P. Aneja, *Voting for Welfare*, 109 CALIF. L. REV. 2013, 2034 (2013).

25. Rosina Lozano, *Vote Aquí Hoy: The 1975 Extension of the Voting Rights Act and the Creation of Language Minorities*, 35 J. POL’Y HIST. 68, 69 (2022).

national origin that favored northern and western Europe,²⁶ significantly increasing the share of immigrants from Asia and Africa.²⁷

The two laws started to transform the United States into a multiracial democracy.²⁸ While European countries accounted for seventy-five percent of the immigrants in the United States in 1960, they accounted for only twelve percent in 2010.²⁹ People of color have grown as a share of the U.S. population from fifteen percent in 1960 to forty-one percent in 2020³⁰ and are expected to make up a majority by 2050.³¹ In 1945, people of color accounted for less than one percent of the members of the U.S. House of Representatives and Senate combined, whereas today they make up over a quarter of members.³² A multiracial coalition (including significant numbers of whites) elected and re-

26. See 8 U.S.C. § 1152(a)(1)(A) (“[N]o person shall receive any preference or priority or be discriminated against in the issuance of an immigrant visa because of the person’s race, sex, nationality, place of birth, or place of residence.”); *id.* § 1152(a)(4)–(5).

27. While the Immigration and Naturalization Act removed some restrictions, it introduced restrictions on immigration from the “Western hemisphere,” which led to a decrease in migration from Latin America. See Kevin R. Johnson, *Fear of an “Alien Nation”: Race, Immigration, and Immigrants*, 7 STAN. L. & POL’Y REV. 111, 112, 119 (1996). As a result, the law was both a victory and a loss for multiracial coalitions, and an example of how gains of some communities of color can come at the expense of other communities of color. See Kevin R. Johnson, *The Beginning of the End: The Immigration Act of 1965 and the Emergence of the Modern U.S.-Mexico Border State*, in THE IMMIGRATION AND NATIONALITY ACT OF 1965: LEGISLATING A NEW AMERICA 116, 121–23 (Gabriel J. Chin & Rose Cuison Villazor eds., 2015).

28. See Tom Ginsburg, *Democratic Backsliding and Multiracial Democracy. A Response to the 2021 Jorde Symposium Lecture by Steven Levitsky*, 110 CALIF. L. REV. 2035, 2038, 2041 (2022).

29. Feysayo Oyolola & Jeanne Batalova, *European Immigrants in the United States*, MIGRATION POL’Y INST. (Jan. 11, 2024), <https://www.migrationpolicy.org/article/european-immigrants-unit-ed-states> [<https://perma.cc/DQG6-2KXV>]; see also Johnson, *supra* note 27 (exploring the history and cultural, socio-economic, and legal effects of the 1965 Immigration Act).

30. William H. Frey, *New 2020 Census Results Show Increased Diversity Countering Decade-Long Declines in America’s White and Youth Populations*, BROOKINGS (Aug. 13, 2021), <https://www.brookings.edu/articles/new-2020-census-results-show-increased-diversity-countering-decade-long-declines-in-americas-white-and-youth-populations> [<https://perma.cc/7RST-GZJ3>].

31. Cheyanne M. Daniels, *Multicultural Americans to Become Majority Population by 2050: Report*, HILL (Jan. 17, 2024, 9:00 AM), <https://thehill.com/homenews/4412311-multicultural-americans-majority-population-2050> (on file with the *Iowa Law Review*). Some of this growth in the share of the United States that is people of color stems from immigration and population growth, but some of it reflects our ability to better acknowledge and recognize multiracial identity. See, e.g., Nicholas Jones, Rachel Marks, Roberto Ramirez & Merarys Ríos-Vargas, *Improved Race and Ethnicity Measures Reveal U.S. Population Is Much More Multiracial*, U.S. CENSUS BUREAU (Aug. 12, 2021), <https://www.census.gov/library/stories/2021/08/improved-race-ethnicity-measures-reveal-united-states-population-much-more-multiracial.html> [<https://perma.cc/BV2M-UHC3>] (indicating that improving question design and coding procedures for the 2020 census reveal the U.S. population is much more multiracial and diverse than previous census measurements).

32. Katherine Schaeffer, *The Changing Face of Congress in 8 Charts*, PEW RSCH. CTR. (Feb. 7, 2023), <https://www.pewresearch.org/short-reads/2023/02/07/the-changing-face-of-congress> [<https://perma.cc/S6BN-GABH>].

elected the nation's first Black president in 2008 and 2012,³³ and public support has increased for diversity, interracial marriage,³⁴ and antidiscrimination laws.³⁵

B. INCREASING RACIAL POLARIZATION AND ANTIDEMOCRATIC ATTITUDES

Despite the transition of the United States toward becoming a well-functioning multiracial democracy, several challenges remain.

For example, as our nation became more racially diverse, our party politics became more racially divided.³⁶ Race is the most significant demographic factor that shapes voting patterns in the United States.³⁷ In 2016, Donald Trump garnered the support of fifty-eight percent of white voters but only eight percent of Black voters (a fifty point gap), and there was a thirty-nine point gap between white voters and both Asian American and Latino voters.³⁸ The gaps were smaller between voters along other binaries—for example, when comparing voters who were straight or LGBT (less than thirty-five points), Protestant or no religion (just over thirty), married or single (fifteen), older or younger (thirteen), male or female (eleven), and high- or low-income (four).³⁹ Similar trends showing the significance of race over other demographic factors are

33. Alex Seitz-Wald, *Obama Had a Coalition. Biden Built a New One and Here's How It's Different*, NBC NEWS (Oct. 30, 2020, 11:50 AM), <https://www.nbcnews.com/politics/2020-election/obama-had-coalition-biden-built-new-one-here-s-how-n1245431> [<https://perma.cc/VUD2-6NR7>].

34. Ginsburg, *supra* note 28, at 2048 (“[I]n 1958, only [four] percent of respondents approved of Black-[w]hite marriages, while in 2021, only [six] percent *dis*-approved.”).

35. Justin McCarthy, *Gallup First Polled on Gay Issues in '77. What Has Changed?*, GALLUP (June 6, 2019), <https://news.gallup.com/poll/258065/gallup-first-polled-gay-issues-changed.aspx> [<https://perma.cc/Y4DM-PKXX>] (describing how support for equal employment opportunities has increased from fifty-six percent in 1977 to ninety-three percent in 2019).

36. *See generally* ZOLTAN L. HAJNAL, DANGEROUSLY DIVIDED: HOW RACE AND CLASS SHAPE WINNING AND LOSING IN AMERICAN POLITICS (2020) (using exit polls and voting data at the mayoral and congressional level to describe this phenomenon).

37. *See id.* at 57–58, 98; Zoltan Hajnal, *How Did We Get Here: Understanding the Demographic Sources of America's Party Divisions*, PROTECT DEMOCRACY (Oct. 10, 2023), <https://protectdemocracy.org/work/understanding-the-demographic-sources-of-americas-party-divisions> [<https://perma.cc/KL2A-S95X>] (showing how race “divides us politically” more than “religion, education, class, and gender” and “as America becomes more diverse, it is also becoming more racially divided in the electoral arena”).

38. *See* Hajnal, *supra* note 37. Communities of color—in particular Asian American and Latino voters with ancestry from varied nations—are not monoliths. *See* Katherine Schaeffer, *Asian Voters in the U.S. Tend to Be Democratic, But Vietnamese American Voters Are an Exception*, PEW RSCH. CTR. (May 25, 2023), <https://www.pewresearch.org/short-reads/2023/05/25/asian-voters-in-the-u-s-tend-to-be-democratic-but-vietnamese-american-voters-are-an-exception> [<https://perma.cc/278C-DSBW>] (showing political diversity among Americans of Chinese, Filipino, Indian, Korean, and Vietnamese ancestry); Jens Manuel Krogstad, Khadijah Edwards & Mark Hugo Lopez, 1. *Hispanics' Views of the U.S. Political Parties*, PEW RSCH. CTR. (Sept. 29, 2022), <https://www.pewresearch.org/2022/09/29/hispanics-views-of-the-u-s-political-parties> [<https://perma.cc/87Z6-Q3P2>] (showing political diversity of Americans of Mexican, Puerto Rican, Cuban, Central American, and South American ancestry).

39. *See* Hajnal, *supra* note 37.

found in elections for other federal and state offices from 1994 to 2006 and 2006 to 2014.⁴⁰

In addition to racial polarization between the political parties, white in-group identification is growing as a political identity. Political scientist Ashley Jardina found that thirty to forty percent of the white population in the United States identify heavily with their in-group as “white.”⁴¹ Jardina found that white identity is “becoming a more salient force in American politics” because many people feel as though they are losing power and status due to demographic changes of the past thirty years stemming from immigration and birth rate differences across racial groups, as well as from symbolic changes like the election of Barack Obama.⁴²

Jardina and other social scientists have found that ethnic antagonism and white racial solidarity correlate with a willingness to abandon democratic norms to maintain status.⁴³ Data show that whites who have a greater degree of racial solidarity back “political leaders who are willing to sidestep democratic norms to advance their agenda.”⁴⁴ They are also more willing to bend rules to “get things done” than “whites who do not possess a racial group consciousness.”⁴⁵ As Ashley Jardina and Robert Mickey wrote, “Some whites’ opposition to democratic principles is rooted, at least in part, in a rejection of racial pluralism; concerns regarding the political claims of racial and ethnic minorities; and the belief that the democratic system works better for people of color, whom they consider less deserving of its benefits.”⁴⁶

The potent combination of racial polarization, cultural anxiety, and antidemocratic attitudes poses unique challenges for multiracial democracy. Recognizing that race is the most significant demographic factor in determining voting behavior, incumbent politicians can strategically entrench their political

40. See HAJNAL, *supra* note 36, at 104–11 (showing the significance of race in various elections from 1996 to 2006); see also Hajnal, *supra* note 37 (showing the same from 2006 to 2014).

41. ASHLEY JARDINA, WHITE IDENTITY POLITICS 8 (2019).

42. Sean Illing, *White Identity Politics Is About More Than Racism*, VOX (Apr. 27, 2019, 11:14 AM), <https://www.vox.com/2019/4/26/18306125/white-identity-politics-trump-racism-ashley-jardina> [https://perma.cc/3CT9-CSBM]; David A. Graham, *Trump’s White Identity Politics Appeals to Two Different Groups*, ATLANTIC (Aug. 8, 2019), <https://www.theatlantic.com/ideas/archive/2019/08/who-does-trumps-white-identity-politics-reach/595189> (on file with the *Iowa Law Review*).

43. See Larry M. Bartels, *Ethnic Antagonism Erodes Republicans’ Commitment to Democracy*, 117 PROC. NAT’L ACAD. SCI. U.S. 22752, 22757 (2020); Jardina & Mickey, *supra* note 13, at 82; Frances E. Lee, *Populism and the American Party System: Opportunities and Constraints*, 18 PERSPS. ON POL. 370, 383 (2020). Granted, most Democratic and Republican voters refrain from punishing candidates in their own party for engaging in antidemocratic behavior. See Matthew Graham & Milan W. Svobik, *Democracy in America? Partisanship, Polarization, and the Robustness of Support for Democracy in the United States*, 114 AM. POL. SCI. REV. 392, 393 (2020) (revealing that only 13.1% of respondents are willing to vote against a candidate in their own party for violating democratic principles).

44. Jardina & Mickey, *supra* note 13, at 80.

45. *Id.* at 79, 85.

46. *Id.* at 81. See generally Andrew Ifedapo Thompson et al., *Anti-Black Political Violence and the Historical Legacy of the Great Replacement Conspiracy*, PERSPS. ON POL., Apr. 9, 2024, at 2 (finding strong associations between demographic change, anti-Blackness, and violence, and how it threatens democratic backsliding).

power by manipulating voting rules or election districts to contain the influence of growing populations of color. They can also rally their declining constituencies with racial dog whistles. Many political operatives have incentives to use these tactics even without AI, and there is no reason to expect they would refrain from deploying a new tool to more effectively implement the tactics. If left unchecked, the widespread availability and use of AI applications can deepen racial polarization, exclude voters of color, and prevent the United States from becoming a well-functioning multiracial democracy.

II. AN OVERVIEW OF AI AND DEMOCRACY

Generally, the term “artificial intelligence” means a field that attempts to make machines intelligent so that they will function appropriately on particular tasks in ways that sometimes match or surpass the capacity of human beings.⁴⁷ AI utilizes “predictions and automation” to mimic “human cognitive functions like problem-solving and learning. . . . and solve complex tasks . . . such as facial and speech recognition, decision-making and translation.”⁴⁸

The first computer program designed to imitate the problem-solving skills of a human was presented at a 1956 Dartmouth conference,⁴⁹ and by 1967, developers had built the first computer based on an artificial neural network (technology that mimics the way the human brain’s neurons signal one another) “that ‘learned’ through trial and error.”⁵⁰

In the 1990s, “machine learning” evolved in which learning algorithms—a set of instructions executed by a computer to learn from data—were developed and trained on historical data so that they could develop solutions through predictions.⁵¹ An algorithm identifies patterns from training data to predict the best outputs, and so an algorithm given “the sandwich contains ___ and jelly” as a prompt would likely complete the sentence with “peanut butter.”⁵² In about 2010, “deep learning” emerged, based on artificial neural networks that utilized greater computational power to categorize data at higher speeds.⁵³

47. Jungherr, *supra* note 8, at 2.

48. *AI vs. Machine Learning vs. Deep Learning vs. Neural Networks: What’s the Difference?*, IBM (July 6, 2023), <https://www.ibm.com/think/topics/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks> [<https://perma.cc/SA6U-D84Z>] [hereinafter *AI vs. Machine Learning*].

49. Rockwell Anyoha, *The History of Artificial Intelligence*, SCI. NEWS (Aug. 28, 2017), <https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence> [<https://perma.cc/C5Z5-D5AU>].

50. Cole Stryker & Eda Kavlakoglu, *What Is Artificial Intelligence (AI)?*, IBM (Aug. 16, 2024), <https://www.ibm.com/topics/artificial-intelligence> [<https://perma.cc/YAL3-ZRA4>].

51. Kristian Lum & Rumman Chowdhury, *What Is an “Algorithm”? It Depends Whom You Ask*, MIT TECH. REV. (Feb. 26, 2021), <https://www.technologyreview.com/2021/02/26/1020007/what-is-an-algorithm> [<https://perma.cc/EXL7-Q6NZ>] (“In statistics and machine learning, we usually think of the algorithm as the set of instructions a computer executes to learn from data.”); RISHI BOMMASANI ET AL., ON THE OPPORTUNITIES AND RISKS OF FOUNDATION MODELS 3 (2022), <https://arxiv.org/pdf/2108.07258> [<https://perma.cc/JRS9-6F7K>].

52. BOMMASANI ET AL., *supra* note 51, at 48.

53. *Id.* at 4; *What Is a Neural Network*, IBM, <https://www.ibm.com/topics/neural-networks> [<https://perma.cc/L7DS-UMXR>].

More recently, generative AI has evolved to be able to create new content, such as text, images, audio, video, and code.⁵⁴ Generative AI models are trained on vast amounts of data selected by humans (e.g., online data) and are able to perform a wide variety of tasks even though they were not specifically trained to do the tasks (e.g., answering a question, writing an essay, taking the bar exam, creating an image or a video, captioning an image, recognizing an object, or determining the sentiment of various emails or social media posts on a topic).⁵⁵ Increasing the size of the training datasets—the “corpora,” or collections of authentic text (e.g., websites, books, news articles, research papers, social media, Wikipedia) as well as audio, pictures, music, code, or other data that the model draws upon to create an output—often increases the power of sequence prediction systems.⁵⁶ The models are also “fine-tuned” on relatively small quantities of task-specific data to mimic particular actions, such as following instructions.⁵⁷

Foundation models are a type of generative AI that can “generate outputs from . . . (prompts) . . . [such as] human language instructions.”⁵⁸ Some foundation models are limited to language (e.g., “large language models” are trained on large amounts of language so that they are able to “understand” natural language instructions from human users and generate language).⁵⁹ The scope of data collected by and outputs produced by some foundation models goes beyond text and includes audio, images, video, computer code, database files, and an understanding of three-dimensional physical shapes.⁶⁰ The output reflects the training dataset. For example, a foundation model designed to create music would have a training dataset of audio songs.⁶¹

54. *What Is Generative AI?*, MCKINSEY & CO. (Apr. 2, 2024), <https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-generative-ai> [<https://perma.cc/87VN-3JXK>].

55. BOMMASANI ET AL., *supra* note 51, at 3.

56. LAURA WEIDINGER ET AL., ETHICAL AND SOCIAL RISKS OF HARM FROM LANGUAGE MODELS 8 (2021), <https://arxiv.org/pdf/2112.04359> [<https://perma.cc/9XKH-SYMV>].

57. JOSH A. GOLDSTEIN ET AL., GENERATIVE LANGUAGE MODELS AND AUTOMATED INFLUENCE OPERATIONS: EMERGING THREATS AND POTENTIAL MITIGATIONS 17 (2023), <https://arxiv.org/pdf/2301.04246.pdf> [<https://perma.cc/C2PF-YPZV>].

58. *What Are Foundation Models?*, AWS, <https://aws.amazon.com/what-is/foundation-models> [<https://perma.cc/JH84-9J5D>].

59. *What Are Large Language Models (LLMs)?*, IBM, <https://www.ibm.com/topics/large-language-models> [<https://perma.cc/2JV9-9KLL>].

60. BOMMASANI ET AL., *supra* note 51, at 7 (“[T]he term language model is simply too narrow for our purpose: as we describe, the scope of foundation models goes well beyond language.” (emphasis omitted)); *id.* at 3 (“A foundation model is any model that is trained on broad data (generally using self-supervision at scale) that can be adapted (e.g., fine-tuned) to a wide range of downstream tasks [F]or example, GPT-3 has 175 billion parameters and can be adapted via natural language prompts to do a passable job on a wide range of tasks despite not being trained explicitly to do many of those tasks.” (citations omitted)).

61. Michael Grothaus, *What Is a “Corpus”? And Why Is Everyone in AI Suddenly Talking About It? Here’s What You Need to Know*, FAST CO. (July 1, 2023), <https://www.fastcompany.com/90916291/what-is-a-corpus-ai-corpora-chatgpt> (on file with the *Iowa Law Review*) (“But a corpus can be made of any kind of data. For example, if you wanted to make an AI that could create music, you would simply include audio songs in its corpus.”).

A *multimodal* foundation model has several types of datasets and can perform several types of tasks. A model that can caption images with text or answer questions about images, for instance, would have been trained on a dataset of both text and images.⁶²

This Article considers generative AI to be a significant advance in *narrow* AI—that is, “data-driven completion of specific tasks in specific domains”⁶³—rather than anticipated *artificial general intelligence* that mimics a human being’s general ability to perform tasks and make decisions across multiple domains, such as cooking a meal, encouraging a dejected work colleague, riding a motorcycle, disciplining a child, or playing piano in a jazz ensemble.⁶⁴ The implications for democracy of artificial general intelligence and other hypothetical advances (such as AI that becomes independent of human control) are speculative at best and beyond the scope of this Article.⁶⁵ Scholars have not yet fully explored, grappled with, or resolved many fundamental challenges of generative AI and related technologies—such as racial harms to democracy. As a result, this Article will focus on anticipating, understanding, and developing principles to mitigate racial harms to democracy that are most likely to arise from existing AI technologies in light of their foreseeable uses and trajectories.

The applications of AI “are nearly endless,” and include chatbots, marketing personalization, facial recognition, natural language processing, and much more.⁶⁶ *Synthetic media machines* enabled by generative AI, for example, can generate text, video, and audio based on input prompts. As detailed below, in the context of democracy these tools can be used to create customized fundraising emails, a campaign commercial, or deepfake video or audio.⁶⁷ Automated *recommender systems* can choose content to present to a user, such as a social media platform’s delivery of ads from candidates touting policing reform to a user’s feed based upon past engagement by the user and similar users with similar content.⁶⁸ Artificial intelligence systems that automatically *classify images* or representations of people are used by social media platforms to detect deepfake videos, as well as by law enforcement to identify activists

62. Jiaqi Wang et al., A Comprehensive Review of Multimodal Large Models: Performance and Challenges Along Different Tasks 5 (Aug. 2, 2024) (unpublished manuscript), <https://arxiv.org/pdf/2408.01319v1> [<https://perma.cc/CM2F-JRUM>].

63. Jungherr, *supra* note 8, at 2.

64. *Id.* (“AI’s current successes do not belong to a still largely fictional AGI but rely instead on advances in narrow AI.”).

65. *See id.* at 9 (“It is important not to be sidetracked by grandiose, but ultimately imaginary, visions of an AGI, but instead focus on specific instances of narrow AI, their inner workings, uses in specific areas of interest, and effects.”).

66. *What’s the Difference Between Data Science and Artificial Intelligence?*, AWS (2024), <https://aws.amazon.com/compare/the-difference-between-data-science-and-ai> [<https://perma.cc/77JA-WMPE>].

67. *Balancing Knowledge and Governance: Foundations for Effective Risk Management of Artificial Intelligence: Hearing Before the Subcomm. on Investigations & Oversight and the Subcomm. on Rsch. & Tech. of the H. Comm. on Sci., Space, & Tech.*, 118th Cong. 2 (2023) (statement of Emily Bender, Professor of Linguistics, Univ. of Wash.) (describing various AI-powered tools).

68. *Id.*

protesting unwarranted police violence or unfair treatment of immigrants.⁶⁹ *Automated decision systems* can be used to flag names that should be removed from voter registration rolls or particular absentee ballot signatures that may be forged (generally jurisdictions require a human to manually review names flagged by automated systems).⁷⁰

Some technologists may consider traditional methods of data science utilized in the context of democracy—such as linear regressions and pattern matching used in some technologies that flag names that should be removed from voter registration rolls—as falling outside of the scope of “artificial intelligence.”⁷¹ The Organization for Economic Co-operation and Development (“OECD”), however, defines an AI system as follows:

An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment.⁷²

This OECD definition of AI—some version of which has been adopted by most governing bodies in the European Union and the United States—is relatively broad.⁷³ The definition would likely include most automated voter roll maintenance and signature matching systems, including “simple pattern matching approaches to detect and flag duplicate registrations,” as well as more

69. *See id.*

70. *See id.*

71. *What’s the Difference Between Data Science and Artificial Intelligence?*, *supra* note 66 (explaining that while both AI and data science use algorithms to analyze data and make predictions, AI goes further and solves “cognitive problems commonly associated with human intelligence, such as learning, pattern recognition, and human-like expression,” and becomes “better at solving problems over time”); Michael Brodie, *Defining Data Science: A New Field of Inquiry* 11 (July 12, 2023) (unpublished manuscript), <https://arxiv.org/abs/2306.16177> [<https://perma.cc/XW2Z-PZXF>] (distinguishing conventional data science from AI-based data science, and noting that AI-based data methods involve machine learning-based methods and deep learning-based methods); Michael Brodie, *A Framework for Understanding Data Science* 27–28 (Feb. 16, 2024) (unpublished manuscript), <https://arxiv.org/pdf/2403.00776> [<https://perma.cc/C6G9-BXS2>] (defining conventional data science methods as “mathematics, simulation, databases, data mining, statistics, probability theory, approximation theory” that are transparent and scrutable (with humans as the learning agents), and defining AI-based methods as “machine learning (ML), evolutionary, heuristic, and generative algorithms” that are “inscrutable lacking solution explanations and results interpretations” (with algorithms as the learning agents)).

72. OECD, REVISED RECOMMENDATION OF THE COUNCIL ON ARTIFICIAL INTELLIGENCE 2–3 (2024), [https://one.oecd.org/document/C/MIN\(2024\)16/FINAL/en/pdf](https://one.oecd.org/document/C/MIN(2024)16/FINAL/en/pdf) [<https://perma.cc/ZG52-3GFG>]; *see also* Stuart Russell, Karine Perset & Marko Grobelnik, *Updates to the OECD’s Definition of an AI System Explained*, OECD.AI POL’Y OBSERVATORY (Nov. 29, 2023), <https://oecd.ai/en/work/ai-system-definition-update> [<https://perma.cc/S5MP-2K4G>] (providing a simplified explanation of an AI system); OECD, EXPLANATORY MEMORANDUM ON THE UPDATED OECD DEFINITION OF AN AI SYSTEM 4 (2024), <https://www.oecd-ilibrary.org/docserver/623da898-en.pdf> [<https://perma.cc/Y6EN-Z8FD>] (providing a more technical explanation of an AI system).

73. *See* Ayesha Gulley & Airlie Hilliard, *Lost in Transl(A)tion: Differing Definitions of AI*, HOLISTIC AI (Feb. 19, 2024), <https://www.holisticai.com/blog/ai-definition-comparison> [<https://perma.cc/FS9N-QPEN>].

advanced machine learning methods used by entities such as the Electronic Registration Information Center (an interstate compact of state election offices that compares voter registration lists to eliminate duplicates).⁷⁴ As a result, this Article analyzes the implications for racially-inclusive democracy of automated election administration systems that maintain voting rolls and match signatures.⁷⁵ While algorithms that can quickly produce millions of different redistricting plans (and may “learn” from previous maps) may be within many conventional legal definitions of AI and have generated literature on the effectiveness of such tools in Voting Rights Act compliance,⁷⁶ redistricting is beyond the scope of this Article.⁷⁷

The various applications of AI to democratic institutions, infrastructure, and election-related communications implicate numerous actors. These include companies that build and train foundation models such as OpenAI, Anthropic, and legacy tech companies (e.g., Amazon, Google, Meta, Microsoft); computational infrastructure companies (e.g., data-driven services that provide storage and computational power, like Amazon Web Services); developers of AI tools that perform specific functions (e.g., surveillance, deepfake detection, signature matching, voice assistance, translation); and search and social media

74. JUNEJA, *supra* note 8, at 15.

75. *See id.* at 11, 15 (relying on OECD’s definition of AI and analyzing automated voter registration maintenance systems); Mekela Panditharatne, *Preparing to Fight AI-Backed Voter Suppression*, BRENNAN CTR. FOR JUST. (Apr. 16, 2024), <https://www.brennancenter.org/our-work/research-report/s/preparing-fight-ai-backed-voter-suppression> [<https://perma.cc/33SX-T97P>] (analyzing the use of automated systems to erroneously identify eligible voters to remove from voter registration lists).

76. *See* Daryl DeFord, Moon Duchin & Justin Solomon, *Recombination: A Family of Markov Chains for Redistricting*, HARV. DATA SCI. REV. 3 (Mar. 31, 2021), <https://hdr.mitpress.mit.edu/pub/1ds8ptxu/release/5> [<https://perma.cc/AG2J-RW7C>] (describing a new form of Markov chain Monte Carlo algorithms that could quickly produce millions of different redistricting plans and explaining the implications of each plan); Jowei Chen & Nicholas O. Stephanopoulos, *The Race-Blind Future of Voting Rights*, 130 YALE L.J. 862, 868 (2021) (in anticipation of conservative judicial rulings scaling back race-consciousness in redistricting, evaluating the effects of “race-blind” algorithms for state legislative redistricting and finding fewer minority opportunity districts and that Republicans would increase their partisan advantage in the South); Moon Duchin & Douglas M. Spencer, *Models, Race, and the Law*, 130 YALE L.J.F. 744, 756 (2021) (finding that “race-blind” algorithms would produce districting plans that significantly underrepresent minority voters relative to their proportion of the population, in part because race is a component of traditional redistricting principles like jurisdictional boundaries); Amariah Becker, Moon Duchin, Dara Gold & Sam Hirsch, *Computational Redistricting and the Voting Rights Act*, 20 ELECTION L.J. 407, 408 (2021) (describing the process of using algorithms to create millions of alternative redistricting maps to identify maps that would increase minority electoral opportunities while also complying with traditional redistricting principles). *But see* Wendy K. Tam Cho & Bruce E. Cain, *Deploying Trustworthy AI in the Courtroom: Lessons from Examining Algorithm Bias in Redistricting AI*, 2023 U. CHI. LEGAL F. 87, 87 n.1 (using “the term AI in an amorphous manner to refer to ‘technology’ writ large (e.g. [sic] software, algorithms, statistical models, optimization heuristics, etc.)” and acknowledging that “the term ‘AI’ has evolved in ways that [the authors] are unable to clearly understand or characterize”).

77. Gerrymandering is relatively simple math that depends largely on the stability of voter preferences and the migration of voters. Currently, the increasing number of U.S. Supreme Court justices who tolerate excessive partisan gerrymandering and vote to scale back the Voting Rights Act is more likely to undermine racially-inclusive democracy than technological developments in redistricting.

platforms. Other relevant players include domestic entities like federal, state, and local candidate and party committees (e.g., Harris for President, Republican National Committee); Super PACs and issue advocacy groups engaging in election-related activities; voting rights and voter mobilization organizations; foreign governments attempting to influence U.S. elections; and individual actors such as community activists and white supremacists. State and local election administrators, legislatures, law enforcement, and other governmental bodies are also using AI-powered tools in ways that shape democracy (e.g., election administration, constituent services, policymaking, translation services, chatbots for voter questions, monitoring political demonstrations). Many of these entities have contractors that play key roles in using automated technologies in democracy, including social media and other communications consultants, data aggregation and analytics firms, and other private vendors.

III. AI'S RACIAL HARMS TO DEMOCRACY

The racial harms of AI to democracy originate not just from deceptive synthetic video and audio (“deepfakes”), but also from numerous technology applications that have been used for years or are emerging but receive little scholarly or media attention. Generative AI will allow hackers to produce code for cyberattacks and text for thousands of frivolous open-record requests targeted to incapacitate local election offices that serve large populations of voters of color. Automated content moderation and deepfake detection systems used by social media platforms, for example, are sometimes less accurate on content created by or featuring people of color. Automated systems that exhibit bias in flagging which voter registrations to purge, which mail-in ballot signatures to invalidate, where to locate polling places, and countless other election administration issues are problematic, even when humans review the outputs and ultimately make the final decisions. An unwarranted perception of mathematical objectivity afforded to some of these applications will often mask racial bias baked into the training dataset and the design of the AI application.⁷⁸

The widespread accessibility of generative AI and other technologies will have a significant sociological impact on democracy.⁷⁹ Advances in natural language processing (a computer’s ability to respond appropriately to text and spoken words), for example, have made foundation models like OpenAI’s

78. See generally *Conversations for Change: Race to the Future? Reimagining the Default Settings of Technology and Society*, NCWIT, <https://ncwit.org/video/race-to-the-future-reimagining-the-default-settings-of-technology-and-society-with-ruha-benjamin-video-playback> [<https://perma.cc/WDG5-NMTQ>] (transcribing conversation with Ruha Benjamin); RUHA BENJAMIN, RACE AFTER TECHNOLOGY: ABOLITIONIST TOOLS FOR THE NEW JIM CODE (2019) (describing a range of discriminatory designs of emerging technologies that foster inequalities); Bruce Schneier, *Ten Ways AI Will Change Democracy*, ASH CTR. FOR DEMOCRATIC GOVERNANCE & INNOVATION (Nov. 6, 2023), <https://ash.harvard.edu/articles/ten-ways-ai-will-change-democracy> [<https://perma.cc/QV7R-MNPW>] (detailing novel ways in which AI might alter how democracy functions beyond AI-generated disinformation).

79. BOMMASANI ET AL., *supra* note 51, at 6 (“[F]oundation model designates a model class that are distinctive in their sociological impact and how they have conferred a broad shift in AI research and deployment.”).

ChatGPT and Google's Gemini broadly available,⁸⁰ and this enables people who are not computer experts to quickly develop large volumes of synthetic text, audio, video, and code.⁸¹ On one hand, this accessibility empowers grassroots activists to engage in more democracy-affirming activism, such as promoting civic engagement and voting in underrepresented Latino, Asian American, and Black communities. On the other hand, the accessibility of the tools will also empower a large group of bad actors who previously lacked the technical expertise, financial resources, or institutional backing to effectively impersonate leaders of color, discourage people of color from voting, bring white supremacist theories into the mainstream, and deploy cyberattacks on county election offices that serve large populations of color.⁸²

If left unchecked, AI can cause real racial harms to American democracy. Democracy requires an active, engaged citizenry,⁸³ and is undermined by the use of AI to engage in deceptive practices and discourage participation by historically-marginalized communities. Civic participation allows people to identify collective challenges and build new coalitions with diverse groups to build solutions,⁸⁴ and the deployment of AI tools to falsely instigate cultural anxiety, racial polarization, and permanent racial factions thwarts democracy.⁸⁵ AI-powered tools that disproportionately erect barriers to political participation for communities of color—whether inadvertently through biased election administration or through concerted efforts like targeted attacks on election offices that serve large populations of color—run counter to democratic principles of equal representation.⁸⁶

Humility is important. The racial harms examined below are not exhaustive. Future technological, racial, and political developments will reveal additional harms and opportunities that we cannot currently foresee,⁸⁷ as well as ways

80. Jim Holdsworth, *What Is NLP (Natural Language Processing)?*, IBM (June 6, 2024), <https://www.ibm.com/topics/natural-language-processing> [<https://perma.cc/8AZD-X6HZ>].

81. West, *supra* note 8 (“AI likely will democratize disinformation by bringing sophisticated tools to the average person. . . . People no longer must be coding experts or video wizards to generate text, images, video, or programs.”).

82. See BOMMASANI ET AL., *supra* note 51, at 45 (“[L]anguage-based foundation models’ ability to take natural language as input, and to generalize to many downstream tasks, could significantly lower the difficulty ‘threshold’ for application development. . . . This could enable even non-ML experts to quickly prototype AI-infused applications.” (citations omitted)).

83. *Whitney v. California*, 274 U.S. 357, 375 (1927) (Brandeis, J., concurring) (“Those who won our independence believed that . . . the greatest menace to freedom is an inert people . . .”).

84. See VERBA ET AL., *supra* note 16, at 110, 112.

85. See Lani Guinier, *No Two Seats: The Elusive Quest for Political Equality*, 77 VA. L. REV. 1413, 1476, 1479 (1991) (explaining that “[t]he documented persistence of racial polarization . . . defeats both of the assumptions supporting the legitimacy of majority rule. . . . [R]acial polarization in the electorate and in the legislative body destroys the reciprocity/virtual representation principle and buries it within racially fixed majorities, thereby transforming majority rule into majority tyranny”).

86. *Reynolds v. Sims*, 377 U.S. 533, 558 (1964) (“The conception of political equality from the Declaration of Independence, to Lincoln’s Gettysburg Address, to the Fifteenth, Seventeenth, and Nineteenth Amendments can mean only one thing—one person, one vote.” (quoting Gray v. Sanders, 372 U.S. 368, 381 (1963))).

87. Bruce Schneier, *Will AI Hack Our Democracy?*, HARV. KENNEDY SCH. MAG., Summer 2023, at 17, 17.

to mitigate some of the harms detailed below. At this juncture, however, considering this myriad of harms together helps advocates, scholars, private industry, and policymakers start to appreciate the potential magnitude and cumulative impact of AI on race and democracy. Examining these disparate applications allows one to identify and categorize common types of harms to democracy and, as detailed in Part IV, develop general principles that can mitigate these and other racial harms to democracy.

A. *RACE AND INFORMATION INTEGRITY*

Artificial intelligence increases the “speed, scale, scope, and sophistication” of racial disinformation and propaganda.⁸⁸ A larger number of people will be empowered to target disinformation about voting and politics at communities of color and disrupt democratic deliberation and decision-making. They will be able to do it more quickly and on a larger scale, with new and more effective tactics, and their messages will be more credible, persuasive, tailored, and difficult to detect as inauthentic.⁸⁹

Widespread use and public awareness of deceptive synthetic content will create new challenges, such as skepticism, disengagement, and callousness about racial justice issues. For example, Black teenagers who witness and record on video the unwarranted killing of their friend by a police officer may face allegations that the video is a deepfake.⁹⁰ They may face greater hurdles in generating support—particularly across racial lines—to hold the officer accountable and enact public policy change.

1. Racial Impersonation and Infiltration

Lower-tech versions of racial impersonation and infiltration of community deliberation provide a roadmap to problems that generative AI will likely make worse.

For example, on Election Day 2016, the operators of the Williams & Kalvin Facebook page—ostensibly two Black men from Atlanta who ran a popular Facebook page focused on Black media and culture—paid for and posted a Facebook ad targeted at Black users. The ad proclaimed: “We don’t have any other choice this time but boycott the election. This time we choose between two racists. No one represents Black people. Don’t go to vote.”⁹¹ After the November 2016 election, an investigation revealed that the Williams & Kalvin

88. Schneier, *supra* note 78.

89. GOLDSTEIN ET AL., *supra* note 57, at 23 tbl.6.

90. See Riana Pfefferkorn, *The Threat Posed by Deepfakes to Marginalized Communities*, BROOKINGS (Apr. 21, 2021), <https://www.brookings.edu/articles/the-threat-posed-by-deepfakes-to-marginalized-communities> [<https://perma.cc/LC7M-FWWX>]; see also Chesney & Citron, *supra* note 9, at 1758, 1785–86 (describing how a “liar’s dividend” phenomenon in “deep fakes make[s] it easier for liars to avoid accountability for things that are in fact true”).

91. YOUNG MIE KIM, UNCOVER: STRATEGIES AND TACTICS OF RUSSIAN INTERFERENCE IN US ELECTIONS: RUSSIAN GROUPS INTERFERED IN ELECTIONS WITH SOPHISTICATED DIGITAL CAMPAIGN STRATEGIES 9 (2018), https://journalism.wisc.edu/wp-content/blogs.dir/41/files/2018/09/Uncover.Kim_v.5.0905181.pdf [<https://perma.cc/65SN-3EWS>].

Facebook account was fake and had been set up and operated by the Russian Internet Research Agency (the “Russian Agency”).⁹²

While African Americans make up just 12.7% of the U.S. population, 37.04% of the unique Facebook pages believed to be created by the Russian Agency were focused on Black audiences.⁹³ In addition, Black audiences were the target of over 38% of the ads purchased by the Russian Agency, accounted for 46.96% of the user impressions, and were the source of 49.84% of the user clicks.⁹⁴

In the 2020 election cycle, domestic political operatives orchestrated a campaign to send robocalls to eighty-five thousand people in largely Black areas. The robocalls—deployed during the height of the COVID-19 pandemic’s social distancing mandates—featured the voice of a woman identifying herself as “Tamika Taylor from Project 1599, the civil rights organization.”⁹⁵ The woman falsely asserted that voting by mail would result in a voter’s personal information being used by police to track persons with outstanding warrants, by credit card companies to collect debts, and by the Centers for Disease Control

92. See Benjamin Fearnow, *Williams & Calvin: Pro-Trump Facebook Stars Reportedly Worked for Kremlin, Accounts Removed*, INT’L BUS. TIMES (Oct. 10, 2017, 1:51 PM), <https://www.ibtimes.com/williams-kalvin-pro-trump-facebook-stars-reportedly-worked-kremlin-accounts-removed-2599559> [<https://perma.cc/X6C8-B9B4>]; Issie Lapowsky, *House Democrats Release 3,500 Russia-Linked Facebook Ads*, WIRED (May 10, 2018, 10:00 AM), <https://www.wired.com/story/house-democrats-release-3500-russia-linked-facebook-ads> [<https://perma.cc/34CC-75S6>]; see also Deen Freelon et al., *Black Trolls Matter: Racial and Ideological Asymmetries in Social Media Disinformation*, 40 SOC. SCI. COMPUT. REV. 560, 573 (2022) (“[T]he [Internet Research Agency] devoted a disproportionate amount of its Facebook advertising budget to microtargeting Black users.” (citing PHILIP N. HOWARD, BHARATH GANESH & DIMITRA LIOTSIU, *THE IRA, SOCIAL MEDIA AND POLITICAL POLARIZATION IN THE UNITED STATES, 2012–2018*, at 3 (2018), <https://demtech.oii.ox.ac.uk/wp-content/uploads/sites/12/2018/12/The-IRA-Social-Media-and-Political-Polarization.pdf> [<https://perma.cc/ZZC7-5G9N>])).

93. See RENEE DIRESTA ET AL., *THE TACTICS & TROPES OF THE INTERNET RESEARCH AGENCY* 12, 21 (2019); ACS *Demographic and Housing Estimates: 2016: ACS 1-Year Estimates Data Profiles*, U.S. CENSUS BUREAU, <https://data.census.gov/table/ACSDP1Y2016.DP05> [<https://perma.cc/5KCP-5UXS>] (indicating a Black population in the United States of 12.7%); see also HOWARD ET AL., *supra* note 92, at 26.

94. See HOWARD ET AL., *supra* note 92, at 23; see also Spencer Overton, *State Power to Regulate Social Media Companies to Prevent Voter Suppression*, 53 U.C. DAVIS L. REV. 1793, 1795–98 (2020) (detailing the 2016 campaign by Russian operatives to target and deceive Black social media users in the United States). In the 2020 election cycle, Russian spending on disinformation targeted at Black voters in the United States continued. See Clarissa Ward, Katie Polgase, Sebastian Shukla, Gianluca Mezzofiore & Tim Luster, *Russian Election Meddling Is Back—via Ghana and Nigeria—and in Your Feeds*, CNN (Apr. 11, 2020, 7:47 AM), <https://www.cnn.com/2020/03/12/world/russia-ghana-troll-farms-2020-ward/index.html> [<https://perma.cc/JA69-Z6LD>].

95. Mychael Schnell, *New York AG James Sues Jacob Wohl, Jack Burkman over Robocalls*, HILL (May 6, 2021, 5:51 PM), <https://thehill.com/homenews/state-watch/552243-new-york-ag-james-sues-jacob-wohl-jack-burkman-over-robocalls> (on file with the *Iowa Law Review*).

and Prevention to “track people for mandatory vaccines.”⁹⁶ Disinformation has also been used to deter voting in Latino communities.⁹⁷

Disinformation in Spanish has also been pervasive in recent elections, in part because platforms have more sophisticated content moderation tools for content in English than in Spanish.⁹⁸ In 2020, for example, YouTube ads in Spanish falsely claiming that Venezuelan dictator Nicolás Maduro supported Democratic presidential candidate Joe Biden were watched over one hundred thousand times in Florida in the nine days just before the 2020 election.⁹⁹ Encrypted messaging applications such as WhatsApp, Telegram, Signal, WeChat, LINE, and Viber are particularly vulnerable to the distribution of non-English political disinformation, as they lack independent fact-checkers and other content moderation structures.¹⁰⁰

Outside of the electoral context, from 2016 to about 2019, a group of domestic non-Black extremists infiltrated a debate within the Black community about #Blaxit (Black people’s exit) and set up fake accounts pretending to be Black users.¹⁰¹ The non-Black infiltrators distributed memes branded in yellow and black designed to mimic Black Lives Matter, created an official Blaxit logo, and took other steps “to create the impression of an emergent movement of African repatriation by a group of Black Americans.”¹⁰² As one infiltrator indicated, “[t]his is like catfishing an entire race.”¹⁰³

96. *Id.*; Nat’l Coal. on Black Civic Participation v. Wohl, 661 F. Supp. 3d 78, 92 (S.D.N.Y. 2023). The Court would later find a violation of the Voting Rights Act because the robocalls were intimidating, threatening, and coercive, and conclude that the conduct was ineligible for First Amendment protection because it constituted a “true threat.” *Id.* at 79, 120.

97. See Panditharatne, *supra* note 75 (“In the 2018 and 2020 elections, operatives targeted Latino communities online by circulating false claims that Immigration and Customs Enforcement (ICE) officers were patrolling voting locations.”).

98. Jude Joffe-Block, *As the 2024 Election Nears, Misinformation Targeting Latinos Gains Attention*, NPR (June 28, 2024, 5:00 AM), <https://www.npr.org/2024/06/25/nx-s1-5013727/new-research-looks-at-how-political-misinformation-is-targeted-at-latinos> [<https://perma.cc/KEV8-DE2T>].

99. Gretel Kahn, *AI, Lies, and Conspiracy Theories: How Latinos Became a Key Target for Misinformation in the U.S. Election*, REUTERS INST. (Mar. 25, 2024), <https://reutersinstitute.politics.ox.ac.uk/news/ai-lies-and-conspiracy-theories-how-latinos-become-key-target-misinformation-use-election> [<https://perma.cc/USST-JM56>]; Gabriel R. Sanchez & Carly Bennett, *Why Spanish Language Mis- and Disinformation Is a Huge Issue in 2022*, BROOKINGS (Nov. 4, 2022), <https://www.brookings.edu/articles/why-spanish-language-mis-and-disinformation-is-a-huge-issue-in-2022> [<https://perma.cc/YT28-6K7V>]; Jeremy B. Merrill & Ryan McCarthy, *Trump Won Florida After Running a False Ad Tying Biden to Venezuelan Socialists*, PROPUBLICA (Nov. 12, 2020, 5:00 AM), <https://www.propublica.org/article/trump-won-florida-after-running-a-false-ad-tying-biden-to-venezuelan-socialists> [<https://perma.cc/R6Z2-ZBPZ>].

100. Katlyn Glover et al., *Encrypted Messaging Applications and Political Messaging: How They Work and Why Understanding Them Is Important for Combating Global Disinformation*, CTR. FOR MEDIA ENGAGEMENT (June 19, 2023), <https://mediaengagement.org/research/encrypted-messaging-applications-and-political-messaging> [<https://perma.cc/87FP-SR5D>].

101. Brandi Collins-Dexter, *Butterfly Attack: Operation Blaxit*, MEDIA MANIPULATION CASEBOOK (Oct. 16, 2020), <https://mediamanipulation.org/case-studies/butterfly-attack-operation-blaxit> [<https://perma.cc/LL9Z-ZG7M>].

102. *Id.*

103. *Id.*

“Digital blackface”—online racial impersonation—is popular in part because of its effectiveness in spreading confusion and discrediting authentic movements. According to one study, presenting as a Black activist is the “most effective predictor of disinformation engagement by far.”¹⁰⁴

While generative AI did not create racial impersonation and is more “an evolution than a revolution in disinformation techniques,”¹⁰⁵ it will build on existing racial cleavages and increase the efficiency and effectiveness of racialized disinformation. Rather than simply rely on social media account still photos and posts written in ethnic vernacular to impersonate people of color, user-friendly and affordable deepfake technology will allow both foreign governments and domestic bad actors to create realistic-sounding synthetic audio and video of people of color—including imitations of trusted voices like current and former elected officials of color, civil rights leaders, celebrities, and other public figures.¹⁰⁶

Before deepfakes and AI-powered text-translation tools, creating disinformation that appeared to come directly from a Spanish-speaking community could require hiring someone who speaks Spanish, but this created real risks that the disinformation scheme would be detected.¹⁰⁷ It also required significant time and resources—one 2017 Russian influence operation that targeted Americans cost \$12.2 million.¹⁰⁸ But AI’s translation abilities allow a disinformation purveyor who speaks no Spanish to quickly create realistic-looking deepfake video and audio and seemingly authentic in-language text with disinformation about election procedures (e.g., the date of the election), candidates, or policy issues.

While the tools will lower the costs of foreign governments and institutional actors to engage in this activity, a disinformation purveyor need not be an arm of the Russian government, a political party, or a SuperPAC. These tools allow lower-resourced organizations and individuals to produce compelling disinformation to influence elections. Such a person can simply be a “lone wolf” with few financial resources who is experiencing cultural anxiety and decides to sow confusion by diminishing the voting strength of Spanish-speaking voters. The low costs to scale the campaign also allow a disinformation purveyor to create more customized deepfakes that are aimed at niche audiences and

104. Freelon et al., *supra* note 92, at 560, 562–63.

105. Christopher Whyte, *Deepfake News: AI-Enabled Disinformation as a Multi-Level Public Policy Challenge*, 5 J. CYBER POL’Y 199, 200 (2020).

106. See, e.g., Bill Wong & Mindy Romero, *Opinion: AI Is Turbocharging Disinformation Attacks on Voters, Especially in Communities of Color*, L.A. TIMES (Mar. 22, 2024, 3:02 AM), <https://www.latimes.com/opinion/story/2024-03-22/artificial-intelligence-disinformation-election-voting> (on file with the *Iowa Law Review*); Matt Brown & David Klepper, *Fake Images Made to Show Trump with Black Supporters Highlight Concerns Around AI and Elections*, L.A. TIMES (Mar. 7, 2024, 11:09 PM), <https://www.latimes.com/world-nation/story/2024-03-08/fake-images-made-to-show-trump-with-black-supporters-highlight-concerns-around-ai-and-elections> (on file with the *Iowa Law Review*).

107. BOMMASANI ET AL., *supra* note 51, at 136.

108. *Id.* at 137.

more persuasive.¹⁰⁹ Deepfake videos that target non-English speakers with disinformation are particularly difficult for platforms to detect and remove.¹¹⁰

Although generative AI models have shortcomings with regard to language (they dominate in English), the models could still allow propagandists who lack direct linguistic and cultural fluency with Black, Latino, Asian American, Arab American, or Tribal communities to more persuasively engage with and pretend to be members of those communities.¹¹¹ Inauthentic social media accounts can sometimes be identified by “a misused idiom” or a “repeated grammatical error,” but generative models trained on the right dataset may produce text, audio, and video that avoid many of these errors.¹¹² AI tools may help those outside of U.S. communities of color overcome basic language and cultural barriers and more accurately use idioms, jargon, colloquialisms, and dialects to more effectively infiltrate the political deliberations of particular communities of color in the United States.¹¹³

The ability to cheaply create large volumes of unique racialized content also adds to the persuasiveness of the disinformation, in part because it seems to come from so many sources. For example, imagine a coordinated disinformation campaign powered by generative AI, launched two days before a closely contested election, that falsely claims a Democratic president and congressional members (including Congressional Black Caucus and Congressional Hispanic Caucus members) favor cuts in federal funding to Historically Black Colleges and Universities (“HBCUs”), Hispanic Serving Institutions, and Tribal Colleges and Universities (“Minority Serving Institutions”). The campaign could be advanced through a false synthetic audio of a meeting of the president,

109. *Id.*

110. Aliya Bhatia, *Election Disinformation in Different Languages Is a Big Problem in the U.S.*, CTR. FOR DEMOCRACY & TECH. (Oct. 18, 2022), <https://cdt.org/insights/election-disinformation-in-different-languages-is-a-big-problem-in-the-us> [<https://perma.cc/KQ4M-JYDA>] (highlighting a study that “found Facebook failed to issue warning labels on [seventy percent] of misinformation in Spanish compared to only [twenty-nine percent] in English”).

111. *Generative AI Is Trained on Just a Few of the World’s 7,000 Languages. Here’s Why That’s a Problem – and What’s Being Done About It*, WORLD ECON. F. (May 17, 2024), <https://www.weforum.org/agenda/2024/05/generative-ai-languages-llm> [<https://perma.cc/6P5W-V9R3>]; see also GOLDSTEIN ET AL., *supra* note 57, at 4.

112. GOLDSTEIN ET AL., *supra* note 57, at 10.

113. Kelly Born, *Will Generative AI Make or Break Democracy?*, PROJECT SYNDICATE (Aug. 10, 2023), <https://www.project-syndicate.org/commentary/generative-ai-could-test-democracies-by-kelly-born-2023-08> (on file with the *Iowa Law Review*) (“Generative AI’s capacity to create persuasive disinformation in multiple languages could also be a boon for foreign adversaries, previously plagued by a lack of language and cultural fluency.”); see also Ellen Barry, *How Russian Trolls Helped Keep the Women’s March Out of Lock Step*, N.Y. TIMES (Sept. 18, 2022), <https://www.nytimes.com/2022/09/18/us/womens-march-russia-trump.html> (on file with the *Iowa Law Review*) (detailing Russian disinformation campaign to undermine American Women’s March by targeting one of its leaders, Palestinian-American Linda Sarsour, by exploiting cultural tensions); Rolf Fredheim & James Pamment, *Assessing the Risks and Opportunities Posed AI-Enhanced Influence Operations on Social Media*, PLACE BRANDING & PUB. DIPL., Feb. 8, 2024, at 7, <https://doi.org/10.1057/s41254-023-00322-5> [<https://perma.cc/2CDU-Z4WU>] (“Our conclusion is that a combination of human operators together with LLM technology does open for new manipulation opportunities. . . . A strengthened capability to produce content en masse in marginal languages is probably the most pressing challenge for diplomatic actors to be aware of.”).

Secretary of the Interior, House and Senate Democratic leaders, and chairs of the Congressional Black Caucus and Congressional Hispanic Caucus during a “confidential” Oval Office discussion. The audio might include remarks about substandard financial controls, high drop-out rates, paltry research output, and inadequate alumni financial support of most Minority Serving Institutions and conclude with a collective agreement to gradually eliminate federal funding of these schools and shift the money to community colleges.

Generative AI allows this synthetic audio to be “substantiated” with, for example: (1) thousands of semantically distinct, outraged social media posts that appear to come from people of color; (2) fake long-form academic research papers purporting to come from different Black political scientists about the failures of Democrats and Congressional Black Caucus Members to support HBCUs since 2000; and (3) fake news reports (all distinct), appearing to come from local news outlets in Texas, Arizona, Nevada, and California, falsely reporting that a particular local congressperson has secretly been conspiring to defund Hispanic Serving Institutions and featuring synthetic video of interviews with local Latino students and alumni criticizing the local congressperson specifically and Democrats nationally for undermining Latino interests.

While lack of public trust in politics is likely to increase as a result of synthetic media generally,¹¹⁴ there are particular implications in the context of race. Since 1960, people of color in the United States have generally had lower levels of trust compared with whites,¹¹⁵ and generative AI poses the risk of deepening distrust of government along racial lines. One 2007 Pew Research Center study, for example, reported that sixty-one percent of Black people and fifty-three percent of Latinos reported low trust, compared with thirty-two percent of whites.¹¹⁶ In 2008, however, the share of white Americans “who say they trust the government to do what is right just about always/most of the time” fell below that of Americans of color, and has remained relatively low.¹¹⁷

This lack of trust facilitated by generative AI is not only likely to affect support for individual candidates, but also civic engagement and participation generally. False claims about alleged voter fraud in communities of color, for

114. Jessica Brandt, *Propaganda, Foreign Interference, and Generative AI*, BROOKINGS (Nov. 8, 2023), <https://www.brookings.edu/articles/propaganda-foreign-interference-and-generative-ai> [https://perma.cc/KK66-F2HG].

115. Sandra Susan Smith, *Race and Trust*, 36 ANN. REV. SOCIO. 453, 456–57 (2010) (observing a consensus that distrust was highest among Black Americans and lowest among whites); Michael Evangelist, *Narrowing Racial Differences in Trust: How Discrimination Shapes Trust in a Racialized Society*, 69 SOC. PROBS. 1109, 1111 (2022). *But see* Cary Wu, Rima Wilkes & David C. Wilson, *Race & Political Trust: Justice as a Unifying Influence on Political Trust*, 151 DAEDALUS J. AM. ACAD. ARTS & SCI. 203, 205–06 (2022) (observing trends that, at times, communities of color have had higher trust in government than whites).

116. PAUL TAYLOR, CARY FUNK & APRIL CLARK, PEW RSCH. CTR., AMERICANS AND SOCIAL TRUST: WHO, WHERE AND WHY 1–2 (2007), <https://assets.pewresearch.org/wp-content/uploads/sites/3/2010/10/SocialTrust.pdf> [https://perma.cc/YKB7-DY4Y]; *see* Tom W. Smith, *Factors Relating to Misanthropy in Contemporary American Society*, 26 SOC. SCI. RSCH. 170, 186–87 (1997).

117. *Public Trust in Government: 1958-2024*, PEW RSCH. CTR. (June 24, 2024), <https://www.pewresearch.org/politics/2024/06/24/public-trust-in-government-1958-2024> [https://perma.cc/7L27-BJCK].

example, can be spread with virality and intensity, can undermine confidence in elections and government, can cause democracy to shrink or collapse, and can enhance racial cleavages and threats of violence.¹¹⁸ The heightened distrust from generative AI also threatens to cause people of all racial backgrounds to discount real injustices like unwarranted police violence, impair community dialogue and interracial coalition building about these issues, and impede the development of strategies, advocacy campaigns, and policy solutions to address problems.¹¹⁹

2. Microtargeting, Manipulative Chatbots, and Cultural Conquest

While generative AI may allow for more precise microtargeting of ads toward communities of color, chatbots trained to persuade, manipulate, and deceive individual users may pose a more significant danger to racially inclusive democracy.

For decades, leading platforms like Meta and Google have developed a business model focused on collecting user data, building a profile on a user, and using the profile to deliver personalized ads to the user.¹²⁰ Data aggregators and data brokers have also long collected and sold data to private companies and governments.

AI can take this to a new level by allowing platforms to analyze microdata collected from a user's reading, purchasing, and other online behavior, then in real time send the user an even more customized message.¹²¹ Because generative AI makes it easy to quickly generate content, it may be even more feasible to make ads customized for a very small group of people.¹²²

Although traditional social media advertisers may have developed a handful of ads for several different audience segments, large language models can create hundreds of ads customized for hundreds of different audience segments.¹²³

118. See generally RICHARD L. HASEN, *CHEAP SPEECH: HOW DISINFORMATION POISONS OUR POLITICS—AND HOW TO CURE IT* (2022) (discussing how American society should respond to the actions of people like former President Trump and his allies, who spread factual falsehoods on social media to undermine the integrity of U.S. elections).

119. See Jeremy Sawyer & Anup Gampa, *Implicit and Explicit Racial Attitudes Changed During Black Lives Matter*, 44 PERSONALITY & SOC. PSYCH. BULL. 1039, 1046 (2018) (explaining that white people became less implicitly pro-white during the Black Lives Matter movement).

120. Jessica Heesen, *AI and Elections: Observations, Analyses and Prospects*, HEINRICH BÖLL STIFTUNG, <https://il.boell.org/en/2022/01/27/ai-and-elections-observations-analyses-and-prospects> [<https://perma.cc/F536-7NKR>] (“Personalized advertising is a legitimate venue for conveying voter information but also a tool for manipulation through microtargeting. . . . This creation of personality profiles is used primarily for personalized advertising and is thus part of the central business model of the dominant digital platforms.”).

121. West, *supra* note 8.

122. Huo Jingnan, *AI-Generated Text Is Hard to Spot. It Could Play a Big Role in the 2024 Campaign*, NPR (June 29, 2023, 5:00 AM), <https://www.npr.org/2023/06/29/1183684732/ai-generate-d-text-is-hard-to-spot-it-could-play-a-big-role-in-the-2024-campaign> [<https://perma.cc/TMV7-UNTR>] (“Realistically, you can’t have someone sit in front of Adobe Premiere and make a video for [one-hundred] people. . . . But generate it with these systems—I think it’s totally possible.” (quoting Alex Stamos)).

123. Fung et al., *supra* note 8 (“Whereas advertisers strategically place a relatively small number of ads, language models such as ChatGPT can generate countless unique messages for you

Currently, campaign messaging often uses simple A/B testing and randomized controlled trials to test which trial messages work best (e.g., determining which of two fundraising solicitation emails sent to two sample groups raises the most money).¹²⁴ Political influence operations also often test possible content on one social media platform to determine which messages to push on other platforms.¹²⁵ As datasets expand, different text, image, audio, and video models evolve, and AI's ability to detect virality and interaction and to test content effectiveness improves,¹²⁶ propagandists will be able to develop content (text, images, video, audio) that is tailored and persuasive to the unique cultural and linguistic knowledge of not only particular ethnic groups but also the intersectional identities of individuals within communities of color.¹²⁷ While an advertiser may have targeted Latinos in Austin, Texas, as a part of a 150,000 person audience in 2016 or 2020, for example, an advertiser will now be able to tailor ads to Latina women in their twenties in Austin who are fans of the FC Dallas Major League Soccer team, have a child in elementary school, and are unaffiliated with a political party.¹²⁸

Although AI machine learning collaboration platforms like Hugging Face currently offer little on creating customized messages for microtargeting,¹²⁹ tailored propaganda may be inevitable if current trendlines continue.¹³⁰ Platforms such as Meta and Google have enjoyed increased revenues as a result of developing tools that allow advertisers to use generative AI to create

personally—and millions for others—over the course of a campaign.”); *see also* Fredheim & Pamment, *supra* note 113, at 3 (“Mass personalisation is another tool that simplifies and accelerates the process of tailoring messaging. Threat actors can exploit LLMs to produce highly targeted material for their campaigns. They can customise prompts to generate material targeted at any language, interest group, or individual. LLMs are ideal tools for targeting and manipulating people’s opinions and beliefs.”).

124. Jack Virag, *A/B Testing and Experimentation in the Obama 2008 and 2012 Campaigns*, STATSIG (Oct. 25, 2022), <https://www.statsig.com/blog/data-experimentation-testing-obama-election-campaigns> [<https://perma.cc/24JM-CQ4Q>]; Arunesh Mathur et al., *Manipulative Tactics Are the Norm in Political Emails: Evidence from 300K Emails from the 2020 US Election Cycle*, BIG DATA & SOC’Y 1 (Jan. 23, 2023), <https://journals.sagepub.com/doi/full/10.1177/20539517221145371> [<https://perma.cc/JG6A-W9UU>].

125. GOLDSTEIN ET AL., *supra* note 57, at 24.

126. Brandt, *supra* note 114 (“Generative AI could increase the persuasiveness of information campaigns by enabling propagandists to test numerous messages at scale before subsequently proliferating those that are most resonant.”).

127. GOLDSTEIN ET AL., *supra* note 57, at 63; Kimberle Crenshaw, *Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics*, 1989 U. CHI. LEGAL F. 139, 140.

128. Sophia Cai, *How AI Is Already Changing the 2024 Election*, AXIOS (Apr. 29, 2023), <https://www.axios.com/2023/04/29/how-ai-already-changing-2024-election> (on file with the *Iowa Law Review*) (“In 2020 or 2016, for example, a 35-year-old male might have been targeted as part of a 250,000-person audience. . . . Now, a campaign can zero in on that voter as one of 25,000 . . .”).

129. *See The AI Community Building the Future*, HUGGING FACE, <https://huggingface.co> [<https://perma.cc/S295-M8ET>] (showing over three thousand sentiment models but nothing on message microtargeting).

130. GOLDSTEIN ET AL., *supra* note 57, at 23 (observing that it is very possible that foundation models will “introduce new forms of deception like tailored propaganda”).

ads tailored to particular audiences, and the platforms are investing more into the technology.¹³¹

Some research reveals that social media ads have limited impact on political behavior,¹³² however, and that using “vast amounts of personal data” to develop even more customized messages will likely not increase the persuasiveness of messages significantly.¹³³ Chatbots may be more effective than microtargeted ads in manipulating users in ways that undermine racially inclusive democracy.

For example, an AI chatbot that was given prompts to “very effectively persuade” conspiracy theorists against their chosen conspiracy reduced conspiracy belief by twenty percent.¹³⁴ This reduction remained among participants two months later, and occurred even among those who had deeply entrenched beliefs.¹³⁵ While the study reduced individual conspiracy beliefs by emphasizing accurate information, it is possible to train an AI chatbot to use misinformation to increase racial distrust, cultural anxiety, and civic disengagement among users.¹³⁶ One examination of Gemini 1.0 models, for example, found the tools moderately persuasive in manipulating a person’s beliefs, and more mature in “persuasion and deception” than in other potentially dangerous areas.¹³⁷

131. Samritha A & Aditya Soni, *Strong Showing by Alphabet, Meta Signal Ad Market Rebound Underway*, REUTERS (Oct. 26, 2023, 12:43 AM), <https://www.reuters.com/technology/strong-ad-show-alphabet-snap-puts-spotlight-meta-2023-10-25> (on file with the *Iowa Law Review*) (“[Meta] is now rolling out tools that use generative AI to create different variations of ad campaigns.”).

132. Katherine Haenschen, *The Conditional Effects of Microtargeted Facebook Advertisements on Voter Turnout*, 45 POL. BEHAV. 1661, 1661–62 (2023) (finding that most Facebook ads do not increase turnout among individuals unlikely to vote, and those that increased turnout were dependent on an alignment of message, audience, and electoral context (1.66 percentage point increase among women in a competitive congressional district who were shown an abortion rights message)).

133. Ben M. Tappin, Chloe Wittenberg, Luke B. Hewitt, Adam J. Berinsky & David G. Rand, *Quantifying the Potential Persuasive Returns to Political Microtargeting*, PROC. NAT’L ACAD. SCI. 1 (June 12, 2023), <https://www.pnas.org/doi/10.1073/pnas.2216261120> [<https://perma.cc/NR8B-EGV5>] (finding that microtargeting political messages generally outperforms other campaign strategies, but that targeting such messages by more than one covariate did not significantly increase the persuasiveness of the message).

134. Thomas H. Costello, Gordon Pennycook & David Rand, *Durably Reducing Conspiracy Beliefs Through Dialogues with AI*, SCI. 1 (Sept. 13, 2024), <https://www.science.org/doi/10.1126/science.adq1814> [<https://perma.cc/N688-3385>] (finding that using the GPT-4 Turbo chatbot to engage in personalized evidence-based dialogues with over two thousand conspiracy theory believers reduced conspiracy belief by roughly twenty percent).

135. *Id.*

136. *See id.* at 7 (“Absent appropriate guardrails, however, it is entirely possible that such models could also convince people to adopt epistemically suspect beliefs . . . or be used as tools of large-scale persuasion more generally.”).

137. Mary Phuong et al., *Evaluating Frontier Models for Dangerous Capabilities*, GOOGLE DEEPMIND 4–9 (Apr. 5, 2024), <https://deepmind.google/research/publications/78150> [<https://perma.cc/77G4-WSQM>].

The ability of AI chatbots to scale up individualized persuasion could “create mass opinion change over short periods of time,” which “could be extremely dangerous, especially when considering nefarious actors.”¹³⁸

The effectiveness of AI chatbots in changing attitudes is due, in part, to their ability to produce several tailored responses and deploy the one that is most persuasive, respond directly to (and refute when necessary) the specific reasoning of a user, and to persistently engage without fatigue at a low cost.¹³⁹ Language models can be trained using an individual’s past speech and text data (e.g., social media posts)¹⁴⁰ to predict future responses from the individual and the types of messages most likely to obtain a desired response from the individual.¹⁴¹ One could envision one-on-one chatbots that are dynamic, interactive, and personalized for an individual.¹⁴²

The forthcoming European Union AI Act bans the use of AI for “[c]ognitive [behavioral] manipulation of people or specific vulnerable groups,”¹⁴³ but such activities are largely unregulated in the United States. To the extent that communities of color disproportionately lack access to “hypernudging” AI tools and are disproportionately vulnerable to manipulation by the tools, AI will deepen racial inequality in democracy.¹⁴⁴

While politicians engaging with (rather than suppressing the votes of) people of color is desirable—and there is not always a bright line dividing legitimate political advocacy from stealth psychological manipulation—such manipulation has unique implications for many communities of color. For people who have internalized experiences like forced assimilation at Indian boarding schools, punishment for speaking Spanish at lunch in the cafeteria at school, employment termination for not straightening curly hair, and

138. Matthew Burtell & Thomas Woodside, *Artificial Influence: An Analysis of AI-Driven Persuasion* 4 (Mar. 15, 2023) (unpublished manuscript), <https://arxiv.org/pdf/2303.08721> [<https://perma.cc/L63J-SBWP>].

139. *Id.*; Costello et al., *supra* note 134, at 2–3.

140. WEIDINGER ET AL., *supra* note 56, at 27.

141. Mike Lewis, Denis Yarats, Yann N. Dauphin, Devin Parikh & Dhruv Batra, *Deal or No Deal? End-to-End Learning for Negotiation Dialogues*, 2017 PROC. CONF. ON EMPIRICAL METHODS NAT. LANGUAGE PROCESSING 2443, 2451 (showing that “it is possible to train dialogue agents end-to-end, but that their ability can be much improved by training and decoding to maximize their goals, rather than likelihood”); GOLDSTEIN ET AL., *supra* note 57, at 24 (“Operators using generative AI models may be able to perform this type of testing at greater scale, which may improve a campaign’s overall impact.”); Panditharatne, *supra* note 75 (“Interactive AI systems can adapt in real time to a voter’s responses; given time and enough input, they might be trained to calculate optimally persuasive arguments tailored to an interlocutor’s positions, or to more accurately predict a voter’s emotional state by analyzing tone or mannerisms.”).

142. GOLDSTEIN ET AL., *supra* note 57, at 2.

143. *EU AI Act: First Regulation on Artificial Intelligence*, EUR. PARLIAMENT (June 18, 2024, 4:29 PM), <https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence> [<https://perma.cc/44Z7-BTAZ>]. As of January 2024, the final act was in the process of being finalized and approved by member states. *Germany to Approve EU’s Planned AI Act*, REUTERS (Jan. 30, 2024, 2:52 PM), <https://www.reuters.com/technology/germany-will-approve-eus-planned-ai-act-sources-2024-01-30> (on file with the *Iowa Law Review*).

144. Thomas Christiano, *Algorithms, Manipulation, and Democracy*, 52 CANADIAN J. PHIL. 109, 109 (2022).

exclusion of their cultural history from school textbooks, manipulation facilitated by generative AI can represent a continuation of unfair cultural conquest that violates values of autonomy,¹⁴⁵ choice, expression, association, and equality that a liberal democracy purports to prioritize.¹⁴⁶ AI applications should not be used to thwart distributed decision-making and self-rule by our increasingly diverse population.¹⁴⁷

3. Fueling Cultural Anxiety

While generative AI can be harnessed to facilitate cross-racial democratic deliberation,¹⁴⁸ propagandists will also use the tools to tap into and increase cultural anxiety, racial polarization, and racial animosity against people of color.¹⁴⁹

Generative AI will allow average people to create content that exacerbates discontent, fears, and anxieties and stokes xenophobia among those who identify as white.¹⁵⁰ Opponents darkened the skin tone of Barack Obama in video ads during the 2008 presidential campaign,¹⁵¹ for example. Generative AI technologies allow individual voters who oppose particular candidates of color to more easily and in larger quantities, do the same to memes, audio, videos, and other content they produce.

145. See WEIDINGER ET AL., *supra* note 56, at 14, 26.

146. Alia Wong, *The Schools That Tried—but Failed—to Make Native Americans Obsolete*, ATLANTIC (Mar. 5, 2019), <https://www.theatlantic.com/education/archive/2019/03/failed-assimilation-native-american-boarding-schools/584017> (on file with the *Iowa Law Review*); Ashley Alese Edwards, *Who Decided Black Hair Is So Offensive Anyway?*, GLAMOUR (Sept. 10, 2020), <https://www.glamour.com/story/black-hair-offensive-timeline> (on file with the *Iowa Law Review*); Joy Sewing, *From Book Bans to Erasing African American Figures from Curriculum, Black History Is Under Attack*, HOUS. CHRON. (Feb. 7, 2023, 9:00 AM), <https://www.houstonchronicle.com/lifestyle/article/black-history-deserves-to-be-told-not-banned-17756596.php> (on file with the *Iowa Law Review*).

147. Jungherr, *supra* note 8, at 3–7 (“AI impacts both the ability of people to achieve self-rule and the perceived superiority of distributed decision-making over expert rule in complex social systems . . .”); see also Iason Gabriel, *Artificial Intelligence, Values, and Alignment*, 30 MIND & MACHS. 411, 425 (2020) (“Designing AI in accordance with a single moral doctrine would . . . involve imposing a set of values and judgments on other people who did not agree with them. For powerful technologies, this quest to encode the true morality could ultimately lead to forms of domination.” (footnote omitted)).

148. See generally H el ene Landemore, *Can AI Bring Deliberative Democracy to the Masses?* (Oct. 9, 2022) (unpublished manuscript), <https://www.law.nyu.edu/sites/default/files/Helen%20Landemore%20Can%20AI%20bring%20deliberative%20democracy%20to%20the%20masses.pdf> [<https://perma.cc/JP6Q-F253>] (exploring how AI can facilitate mass deliberation).

149. See GOLDSTEIN ET AL., *supra* note 57, at 14 (“Preexisting polarization and fragmentation . . . can give influence operators a foothold to tailor their messaging to narrower audiences, sow division, and degrade social capital and institutional trust.”).

150. West, *supra* note 8 (“Generative AI can develop messages aimed at those upset with immigration . . . [or] critical race theory . . .”).

151. Max Ehrenfreund, *Obama’s Skin Looks a Little Different in These GOP Campaign Ads*, WASH. POST (Dec. 29, 2015, 6:01 AM), <https://www.washingtonpost.com/news/wonk/wp/2015/12/29/obamas-skin-looks-a-little-different-in-these-gop-campaign-ads> (on file with the *Iowa Law Review*); Solomon Messing, Maria Jabon & Ethan Plaut, *Bias in the Flesh: Skin Complexion and Stereotype Consistency in Political Campaigns*, 80 PUB. OP. Q. 44, 45–46 (2016).

A majority of white supremacists are now unaffiliated individuals rather than members of organized hate groups.¹⁵² The accessibility of generative AI tools empowers these individuals to quickly create massive amounts of unique and high-quality text, audio, and video promoting their ideology without significant financial resources or technical expertise.¹⁵³ In the past, white supremacists have been early adopters of new technologies, including film,¹⁵⁴ radio,¹⁵⁵ computer bulletin boards and websites,¹⁵⁶ social media,¹⁵⁷ and streaming video,¹⁵⁸ and we should anticipate their early adoption of generative AI tools as well. Recognizing AI's capacity to produce and distribute propaganda quickly, white supremacists may use such tools to recruit and radicalize adherents, create videos and manifestos (including those depicting violence against people of color and derogatory racial stereotypes), fundraise, and promote their principles of white cultural and genetic superiority.¹⁵⁹

Generative AI will also be a particularly valuable tool for a specific subset of white supremacists who attempt to appear moderate so that they may infiltrate traditional political institutions and insert white nationalist priorities—such as opposition to immigration and multicultural programs—into mainstream policy debates.¹⁶⁰ White supremacists use extreme platforms to discuss and develop strategy, and then use “sock puppet” accounts on platforms such as Facebook and YouTube to execute their strategy to bring their ideas into the mainstream. As the Southern Poverty Law Center explains:

152. Daniel Byman, *When Hate Goes Viral*, FOREIGN POL'Y (Mar. 23, 2022), <https://foreignpolicy.com/2022/03/23/white-supremacist-terrorism-social-media-internet> [<https://perma.cc/ZE7P-XJ2C>]; see also *With Hate in Their Hearts: The State of White Supremacy in the United States*, ADL (Mar. 3, 2017), <https://www.adl.org/resources/report/hate-their-hearts-state-white-supremacy-united-states> [<https://perma.cc/R87N-XD7X>] (“Most white supremacists do not belong to organized hate groups, but rather participate in the white supremacist movement as unaffiliated individuals.”).

153. GOLDSTEIN ET AL., *supra* note 57, at 1; West, *supra* note 8.

154. See *100 Years Later, What's the Legacy of 'Birth of a Nation'?*, NPR (Feb. 8, 2015, 5:56 PM), <https://www.npr.org/sections/codeswitch/2015/02/08/383279630/100-years-later-whats-the-legacy-of-birth-of-a-nation> [<https://perma.cc/7T4S-XMVN>].

155. See Allison C. Meier, *An Affordable Radio Brought Nazi Propaganda Home*, JSTOR DAILY (Aug. 30, 2018), <https://daily.jstor.org/an-affordable-radio-brought-nazi-propaganda-home> [<https://perma.cc/U8GA-qJT9>].

156. HEATHER J. WILLIAMS, ALEXANDRA T. EVANS, JAMIE RYAN, ERIC E. MUELLER & BRYCE DOWNING, *THE ONLINE EXTREMIST ECOSYSTEM* 5 (2021) (“White supremacists’ adoption of online platforms began near the founding of computer networks in the early 1980s, when white power activists . . . established public bulletin board systems (BBSs).”).

157. *Id.* at 6 (“As social media sites gained prominence and attracted larger user bases over the early and mid-2000s, white power activists, bands, and other groups joined mainstream platforms . . .”).

158. Chloe Mayer, *Facebook, Twitter Battling Buffalo Shooting Video 48 Hours on*, NEWSWEEK (May 16, 2022, 11:43 AM), <https://www.newsweek.com/buffalo-shooting-video-online-facebook-twitter-social-media-response-1706961> [<https://perma.cc/6DDA-2JAR>].

159. See Byman, *supra* note 152; ANTI-DEFAMATION LEAGUE, *WITH HATE IN THEIR HEARTS: THE STATE OF WHITE SUPREMACY IN THE UNITED STATES* 17 (2017), <https://www.adl.org/sites/default/files/state-of-white-supremacy-united-states-2015.pdf> [<https://perma.cc/Q8MX-6HG3>].

160. *White Nationalist*, S. POVERTY L. CTR., <https://www.splcenter.org/fighting-hate/extremist-files/ideology/white-nationalist> [<https://perma.cc/NK3T-VQ2V>] (“Mainstreamers believe that infiltrating and subverting the existing political institutions is the only realistic path to power.”).

The continued radicalization of the GOP has greatly aided the white nationalist movement, exhibited by the party's embrace of such racist concepts as the "great replacement," vilification of immigrants, attacks on reproductive care, and demonization of queer and trans people. White nationalists will continue to abet the broader right's attacks on marginalized people and communities through propaganda production, participation in protests and other forms of intimidation and even violence. Twitter's choice to reinstate extremists and slacken enforcement of hate speech policies will mean that more people will be exposed to white nationalist propaganda and harassment.¹⁶¹

B. AI MODEL DESIGN CAN ENTRENCH RACIAL HIERARCHY

Even absent intentional attempts to use AI technologies to engage in racially targeted voter deception or to stoke cultural anxiety, the architecture of AI foundation models alone can fortify political, cultural, language, and racial outlooks of the past. Bias embedded in training datasets and the design of AI used to create content, moderate content, and detect deepfakes can replicate and embed racial hierarchy in our democratic debates.

Because foundation models are optimized to reflect their training data as accurately as possible to detect patterns, we should expect them to "reinforce stereotypes and unfair discrimination by default."¹⁶² Training datasets may be biased because they are gathered from contexts that reflect societal inequality and/or because some communities are better represented in training data (an AI image generator trained on portraits and photos of the forty-four white males and one Black male who have served as U.S. presidents, for example, would likely not produce an image of a Black female U.S. president).¹⁶³

In addition to problematic data, biases arise from the design of the AI model,¹⁶⁴ the lack of diversity of those who develop models,¹⁶⁵ and when adapting the foundation model to particular applications, such as text summarization or the creation of images.¹⁶⁶ While guardrails may help reduce direct racial bias, many racial indicators are implied rather than explicitly stated, and research has found biases are therefore embedded that reflect

161. *Id.*

162. WEIDINGER ET AL., *supra* note 56, at 11.

163. *Id.*; Tom B. Brown et al., *Language Models Are Few-Shot Learners*, 34 CONF. ON NEURAL INFO. PROCESSING SYS. 1877, 1889 (2020); *see also* VIRGINIA EUBANKS, AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR (2018) (explaining how marginalized groups are subjected to "collective red-flagging" and "feedback loop of justice"); CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY 86–87 (2017) (demonstrating how predictive crime models for certain neighborhoods can be populated with "nuisance" crimes to "creat[ing] a pernicious feedback loop . . . and justif[ying] more policing").

164. BOMMASANI ET AL., *supra* note 51, at 130.

165. *Id.* at 133 ("As with all algorithmic systems, poor representation and diversity of stakeholders and marginalized communities in decision-making bodies that develop or apply foundation models is inherently problematic, and may contribute to greater experienced harm for these communities.")

166. *Id.* at 130.

both the “training data and societal biases at large.”¹⁶⁷ Racial indicators are often implied (e.g., educational institutions attended, social media connections, musical preferences), and it is difficult to know when AI makes an inference (even when systems are designed to disregard race). These biases result in skewed outputs—such as representational bias, performance disparities (e.g., higher error rates in detecting deepfakes involving people with darker skin), and stereotypes.¹⁶⁸ While many Americans aspire to break with past patterns of discrimination, “AI-based predictions and classifications based on past patterns risk replicating systemic inequalities and even structural discrimination.”¹⁶⁹

In the context of democracy, biases within the foundation models have the potential to reproduce and entrench racial assumptions in our politics and disparities in the allocation of political power. Several studies, for example, reveal that language models reflect a consistent bias against Black Americans.¹⁷⁰ As Professor Andreas Jungherr wrote:

People’s visibility to AI depends on their past representation in data. AI has trouble recognizing those who belong to groups underrepresented in the data used to train it This general pattern is highly relevant to democracy: for example, the systematic invisibility of specific groups means they would be diminished in any AI-based representation of the body politic and in predictions about its behavior, interests, attitudes, and grievances. Accordingly, already disenfranchised people could risk further disenfranchisement and discrimination in the roll out of government services, the development of policy agendas based on digitally mediated preferences and voice, or face heightened persecution from the state security apparatus AI might contribute to an increase of resources for the already

167. Melissa Warr, Nicole Jakubczyk Oster & Roger Isaac, *Implicit Bias in Large Language Models: Experimental Proof and Implications for Education*, J. RSCH. ON TECH. EDUC. 1 (Aug. 28, 2024), <https://www.tandfonline.com/doi/full/10.1080/15391523.2024.2395295> (on file with the *Iowa Law Review*).

168. BOMMASANI ET AL., *supra* note 51, at 130–31; WEIDINGER ET AL., *supra* note 56, at 9 (“Perpetuating harmful stereotypes and discrimination is a well-documented harm in machine learning models that represent natural language.”); Amit Haim, Alejandro Salinas & Julian Nyarko, *What’s in a Name? Auditing Large Language Models for Race and Gender Bias* 9–10 (Feb. 29, 2024) (unpublished manuscript), <https://arxiv.org/pdf/2402.14875> [<https://perma.cc/5W5N-VXWL>] (investigating the biases of GPT-4 and other language models, and finding that the advice given systematically disadvantages people of color and women, and that names associated with Black women receive the least advantageous outcomes).

169. Jungherr, *supra* note 8, at 3.

170. Jwala Dhamala et al., *BOLD: Dataset and Metrics for Measuring Biases in Open-Ended Language Generation*, 2021 ACM CONF. ON FAIRNESS ACCOUNTABILITY & TRANSPARENCY 862, 868 (finding that “the biases in three different LMs and a comparison with Wikipedia texts show that LMs are prone to more frequently generating texts with negative connotations towards a particular group of people or an idea than others,” especially demonstrating a consistent bias against African Americans); Moin Nadeem, Anna Bethke & Siva Reddy, *StereoSet: Measuring Stereotypical Bias in Pretrained Language Models*, 59 PROC. ASS’N FOR COMPUTATIONAL LINGUISTICS 5356, 5356 (2021) (“Since pretrained representations are obtained from learning on massive text corpora, there is a danger that stereotypical biases in the real world are reflected in these models.”).

privileged by making their voices, interests, attitudes, concerns, and grievances more visible and accessible to decision-makers.¹⁷¹

Artificial intelligence generally “can compound existing inequities by producing unfair outcomes, entrenching systems of power, and disproportionately distributing negative consequences of technology to those already marginalized.”¹⁷² These tools can affect the electoral process and in turn reproduce and even deepen current inequality in voter participation, political representation, and policymaking process through government distribution of tax benefits, expenditures, and other opportunities.

1. AI Models Can Minimize the Perspectives of People of Color

Foundation models trained on datasets that underrepresent people of color can lock in and magnify the frameworks, language, and political perspectives of a shrinking share of the population.¹⁷³ In scraping the web for pictures and text and accepting them as “representative,” an AI foundation model will perpetuate the viewpoints that are dominant of those who post on the web and risk “increasing power imbalances, and further reifying inequality.”¹⁷⁴ For example, BERT, a language model introduced in 2018 by Google researchers

171. Jungherr, *supra* note 8, at 7.

172. BOMMASANI ET AL., *supra* note 51, at 130; *see also* Ngozi Okidegbe, *To Democratize Algorithms*, 69 UCLA L. REV. 1688, 1710–11 (2023) (addressing “how algorithmic use operates to entrench and legitimate state practices that suppress the democratic participation of these groups, reinforcing their oppression and their political powerlessness in governance”); *see also* BENJAMIN, *supra* note 78, at 87–90; Anjalie Field, Su Lin Blodgett, Zeerak Waseem & Yulia Tsvetkov, *A Survey of Race, Racism, and Anti-Racism in NLP*, 59 PROC. ANN. MEETING ASS’N FOR COMPUTATIONAL LINGUISTICS 1905, 1907–13 (2021); Timnit Gebru, *Race and Gender*, in THE OXFORD HANDBOOK OF ETHICS OF AI 252, 256–57 (Markus Dirk Dubber, Frank Pasquale & Sunit Das eds., 2021) (explaining how AI and data-driven decision-making creates negative feedback loops); Alex Hanna, Emily Denton, Andrew Smart & Jamila Smith-Loud, *Towards a Critical Race Methodology in Algorithmic Fairness*, FAT*’20: PROC. CONF. ON FAIRNESS ACCOUNTABILITY & TRANSPARENCY 501, 501–10 (2019).

173. *See* WEIDINGER ET AL., *supra* note 56, at 14 (“A LM trained on language data at a particular moment in time risks not just excluding some groups, but also enshrining temporary values and norms without the capacity to update the technology as society develops.”). It is also possible that while underrepresenting less affluent elements of racially diverse populations, some large language models also underrepresent older, less racially diverse populations whose perspectives may be underrepresented in online text and whose share of the U.S. population may be much greater than their share of the AI workforce. *See* Shibani Santurkar et al., *Whose Opinions Do Language Models Reflect?*, 40 PROC. INT’L CONF. ON MACH. LEARNING 29971, 29980–84 (2023).

174. Emily M. Bender, Timnit Gebru, Angelina McMillan-Major & Shmargaret Schmittell, *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?*, FACCT 614 (Mar. 2021), <https://doi.org/10.1145/3442188.3445922> [<https://perma.cc/V8DA-GK2C>] (“In accepting large amounts of web text as ‘representative’ of ‘all’ of humanity we risk perpetuating dominant viewpoints, increasing power imbalances, and further reifying inequality.”). Racial underrepresentation in online materials may stem from disparities in digital access. *See* Jessica Douglas, *Bridging the Digital Divide in Indian Country*, HIGH COUNTRY NEWS (Feb. 4, 2021) (citing H. TROSTLE, BUILDING INDIGENOUS FUTURE ZONES: FOUR TRIBAL BROADBAND CASE STUDIES 4 (Feb. 2021)), <https://www.hcn.org/issues/53-4/indigenous-affairs-technology-bridging-the-digital-divide-in-indian-country> [<https://perma.cc/8KNY-8WAQ>] (showing that only approximately sixty percent of Tribal lands in the continental United States had high-speed Internet access).

that made major advancements in natural language processing,¹⁷⁵ “appears to encode an Anglocentric perspective by *default*, which can amplify majority voices and contribute to *homogenization* of perspectives or monoculture.”¹⁷⁶ Another study found the perspectives of several language models were most aligned with lower income, moderate, Protestant, or Roman Catholic groups (possibly because the models were trained on snapshots of the internet),¹⁷⁷ while finding OpenAI models were more aligned with liberal, high-income, well-educated, and not religious groups other than Buddhists, Muslims, or Hindus (that is, aligning with the “predominantly young Southeast Asian and White with a college degree” workforce provided by OpenAI in disclosing its workplace diversity statistics).¹⁷⁸

System design and societal bias can result in chatbots that produce distorted portrayals of people of color and their political movements. For example, to the extent that traditional news outlets produce stories about Black Lives Matter protests that underreport typical peaceful demonstrations and overcover isolated dramatic confrontations with police,¹⁷⁹ a primary source of online data used to train foundation models is likely skewed. The models are much more likely to produce content that minimizes legitimate claims for police accountability and inflate the extent to which the movement represents disorder and violence.¹⁸⁰ The phenomenon may produce a feedback loop in which human journalists, reporters, editors, and producers use the racially skewed insights produced by AI tools to frame new stories they develop, thereby further embedding the bias in future training datasets.¹⁸¹

Training data, system designers, and other factors help determine whose perspectives and narratives a language model will promote, and when the perspectives, narratives, and experiences of communities of color are underrepresented they are marginalized.¹⁸² Language is essential in how we interact with each other, express our differences, find common ground, and engage in democracy. Technologies that skew this exchange along racial lines not only undermine fundamental rights of communities of color but also the

175. See Britney Muller, *BERT 101 - State of the Art NLP Model Explained*, HUGGING FACE (Mar. 2, 2022), <https://huggingface.co/blog/bert-101> [<https://perma.cc/gTL2-B7KF>].

176. BOMMASANI ET AL., *supra* note 51, at 131.

177. Santurkar et al., *supra* note 173, at 8.

178. *Id.* at 8–9.

179. Kathleen Bartzan Culver & Douglas M. McLeod, “Anti-Riot” or “Anti-Protest” Legislation? *Black Lives Matter, News Framing, and the Protest Paradigm*, 4 JOURNALISM & MEDIA 216, 225–27 (2023) (finding that the most frequent news frame in covering state legislation to control Black Lives Matter protests was fighting crime with less attention to free expression and race).

180. See Bender et al., *supra* note 174.

181. See WEIDINGER ET AL., *supra* note 56, at 14.

182. *Id.* at 24 (“The choice of training corpora is thus an important one: it affects whose perspectives will be better represented by LM predictions, and whose narratives the LM promotes. . . . Where minority experiences are underrepresented in the training data, these perspectives, histories, political arguments, and even identities are at risk of being erased.” (emphasis omitted)).

legitimacy of democracy broadly.¹⁸³ This slows societal progress toward a truly multiracial democracy as the United States becomes more racially diverse.¹⁸⁴

These issues are not simply limited to the *bias* of foundation models but also stem from the foundation models' limitations in *adequately serving all* in a pluralistic, diverse society.¹⁸⁵ Because foundation models typically answer questions in mathematical ways by defaulting to averages or dominant patterns,¹⁸⁶ they often fail to adequately serve diverse perspectives—including those of communities of color in the context of democracy. Technologists have only started to acknowledge these limitations of foundation models and are trying to develop technical solutions.¹⁸⁷

2. Language Models Can Suppress Language Minority Participation

The bias for English in large language models (text) and other foundation models (e.g., audio) could shape various aspects of U.S. democracy, including entrenching the cultural domination of the past and obstructing diversity moving forward—in language, dialect, norms, and political perspectives.¹⁸⁸ This bias could adversely affect communities of Asian, American Indian, Alaska Native, and Spanish heritage who are protected by language assistance provisions

183. ELEANOR SHEARER, SABRINA MARTIN, ANDRÉ PETHERAM & RICHARD STIRLING, RACIAL BIAS IN NATURAL LANGUAGE PROCESSING 2 (2019), https://ec.europa.eu/futurium/en/system/files/ged/shared_racial_bias_in_natural_language_processing_-_report.pdf [<https://perma.cc/M7P3-N64J>].

184. Bender et al., *supra* note 174, at 614 (“In accepting large amounts of web text as ‘representative’ of ‘all’ of humanity we risk perpetuating dominant viewpoints, increasing power imbalances, and further reifying inequality.”).

185. Sorensen et al., *supra* note 5, at 46280, 46285 (recognizing that foundation models should serve “a diverse set of human values and perspectives” and fail to do so, and that there are open questions on whether “a system be pluralistic?”); see also Shangbin Feng, Chan Young Park, Yuhan Liu & Yulia Tsvetkov, *From Pretraining Data to Language Models to Downstream Tasks: Tracking the Trails of Political Biases Leading to Unfair NLP Models*, 61 PROC. ANN. MEETING ASS’N FOR COMPUTATIONAL LINGUISTICS 11737, 11743 (2023) (finding that “for hate speech detection, models with left-leaning biases exhibit better performance towards hate speech directed at widely-regarded minority groups”).

186. Gordon et al., *supra* note 5, at 2; *Value Kaleidoscope*, *supra* note 5, at 19938 (asserting that “contemporary supervised AI systems primarily wash out variation by aggregating opinions or preferences with majority votes”). When foundation models address diversity, they sometimes do it in ways that invite public scrutiny. See, e.g., Casey Newton, *Google Hits Pause on Gemini’s People Pictures*, PLATFORMER (Feb. 22, 2024), <https://www.platformer.news/google-gemini-ai-photos-diversity-pause> [<https://perma.cc/gWRA-M5TY>] (reporting on Google’s Gemini chatbot creation of racially diverse Nazis).

187. Sorensen et al., *supra* note 5, at 46280 (proposing a roadmap to allow foundation models to serve “a diverse set of human values and perspectives”); Gordon et al., *supra* note 5, at 1, 2 (introducing “jury learning, a supervised learning architecture that closes th[e] gap through the metaphor of a jury”); *Value Kaleidoscope*, *supra* note 5, at 19938 (constructing a dataset with “218k contextualized values, rights, and duties . . . connected to 31k human-written real-life situations” that attempts to “assess[] the relevance and valence . . . of contextualized pluralistic human values, rights, and duties”).

188. Paresh Dave, *ChatGPT Is Cutting Non-English Languages Out of the AI Revolution*, WIRED (May 31, 2023, 7:00 AM), <https://www.wired.com/story/chatgpt-non-english-languages-ai-revolution> [<https://perma.cc/Z7TS-QF8P>].

of the Voting Rights Act,¹⁸⁹ as well as other people of color who speak other non-English languages (e.g., Haitian-Creole, Arabic) and “non-standard” dialects of English.

In the United States, race is connected to language.¹⁹⁰ While people of color account for just over forty percent of the U.S. population, they account for over eighty-six percent of the limited English-proficient population in the United States.¹⁹¹ Of the total limited English-proficient population in the United States, the largest shares speak Spanish (sixty-four percent), followed by Chinese (six percent), Vietnamese (three percent), Korean (two percent), and Tagalog (two percent).¹⁹²

AI tools can reduce language barriers and allow Americans with limited-English proficiency to more effectively participate in democracy. These tools can help with the translation of candidate names and party designations on ballots and the translation of and explanation of complex ballot initiatives and referenda.¹⁹³ They can help facilitate communication not just of speakers of relatively popular languages like Spanish, but also speakers of Tribal languages and other less popular languages.¹⁹⁴ The natural language processing capabilities of language models also allow them to facilitate democratic participation of those who primarily speak languages without written characters and populations with low-literacy rates. Various AI-powered tools can also help increase democratic engagement outside of the voting booth, including translation tools, voice assistance tools, chatbots, content moderation tools, recommendation algorithms, search functions, and news aggregation tools.

These tools may fall short in serving limited English-proficient communities, however, because a very small share of the world’s over six thousand languages are represented in foundation models.¹⁹⁵ English is by far the most dominant

189. 52 U.S.C. § 10503 (b) & (c) (requiring bilingual election materials in jurisdictions with large numbers or a share of a single language minority who are limited-English proficient).

190. See generally RACIOLINGUISTICS: HOW LANGUAGE SHAPES OUR IDEAS ABOUT RACE (H. Samy Alim, John R. Rickford & Arnetta F. Ball eds., 2016) (examining the role of language in shaping race).

191. Jie Zong & Jeanne Batalova, *The Limited English Proficient Population in the United States in 2013*, MIGRATION POL’Y INST. (July 8, 2015), <https://www.migrationpolicy.org/article/limited-english-proficient-population-united-states-2013> [<https://perma.cc/B2W7-VPAC>].

192. *Id.*

193. See Yonathan A. Arbel & Shmuel I. Becher, *Contracts in the Age of Smart Readers*, 90 GEO. WASH. L. REV. 83, 95–99 (2022) (discussing the ability of language models to read, analyze, and assess contracts, disclosures, and other complex language).

194. See Séamus Lankford, Haithem Afli & Andy Way, *adaptMLLM: Fine-Tuning Multilingual Language Models on Low-Resource Languages with Integrated LLM Playgrounds*, INFORMATION 2 (Nov. 29, 2023), <https://www.mdpi.com/2078-2489/14/12/638> [<https://perma.cc/3EA5-G3TN>]; Xuan-Phi Nguyen, Sharifah Mahani Aljunied, Shafiq Joty & Lidong Bing, *Democratizing LLMs for Low-Resource Languages by Leveraging Their English Dominant Abilities with Linguistically-Diverse Prompts*, 62 PROC. ANN. MEETING ASS’N FOR COMPUTATIONAL LINGUISTICS, 3501, 3501 (2024); Vivek Iyer et al., *Exploring Very Low-Resource Translation with LLMs: The University of Edinburgh’s Submission to AmericasNLP 2024 Translation Task*, 2024 PROC. 4TH WORKSHOP ON NAT. LANGUAGE PROCESSING FOR INDIGENOUS LANGUAGES AMS. 209, 209.

195. BOMMASANI ET AL., *supra* note 51, at 23.

language.¹⁹⁶ Although Spanish is the second most popular language in the United States¹⁹⁷ and is represented in large language models, it is less well represented in the datasets used to train large language models than English. Native American, Alaska Native, Filipino, Cambodian, Vietnamese, Hmong, and other languages may either be unrepresented or underrepresented, due in part to the lack of datasets for training. A related issue is that a “shocking” amount of non-English content on the web has been machine-translated from English, and as a result many large language models have been trained on low-quality examples of non-English languages.¹⁹⁸

Applications built on foundation models may also struggle with language variation—including dialects of English spoken by some African Americans, Latinos, and Asian Americans. One 2017 study showed the failures of natural language processing (a computer’s ability to “understand” text and spoken words) for English spoken by some African Americans.¹⁹⁹

As a result, tools powered by AI that facilitate democratic engagement and participation by many people of color in the United States—including translation tools, voice assistance tools, chatbots, content moderation tools, recommendation algorithms, search functions, and news aggregation tools—may be less accurate, nuanced, and effective in languages other than standard English.²⁰⁰ As indicated in ChatGPT-4’s report card, a “majority of pretraining data and our alignment data is in English,”²⁰¹ and as a company staffer remarked on a support forum, “[a]ny good Spanish results are a bonus.”²⁰²

196. *Id.* at 24. As of 2020, the multilingual foundation models that have been developed were trained on around one-hundred languages. *Id.*; see also WEIDINGER ET AL., *supra* note 56, at 17; Emily M. Bender, *The #BenderRule: On Naming the Languages We Study and Why It Matters*, GRADIENT (Sept. 14, 2019), <https://thegradient.pub/the-benderrule-on-naming-the-languages-we-study-and-why-it-matters> [<https://perma.cc/DWF2-N722>].

197. Sandy Dietrich & Erik Hernandez, *Nearly 68 Million People Spoke a Language Other Than English at Home in 2019*, U.S. CENSUS BUREAU (Dec. 6, 2022), <https://www.census.gov/library/stories/2022/12/languages-we-speak-in-united-states.html> [<https://perma.cc/RZ29-HTPW>].

198. Brian Thompson, Mehak Preet Dhaliwal, Peter Frisch, Tobias Domhan & Marcello Federico, *A Shocking Amount of the Web Is Machine Translated: Insights from Multi-Way Parallelism*, 2024 FINDINGS ASSOC. FOR COMPUTATIONAL LINGUISTICS 1763, 1763–68. Bias is also difficult to evaluate with multilingual language models. Zeerak Talat et al., *You Reap What You Sow: On the Challenges of Bias Evaluation Under Multilingual Settings*, 2022 PROC. BIGSCIENCE EPISODE #5, at 26, 26–33, (explaining that the challenges of evaluating bias with large multilingual language models).

199. Su Lin Blodgett & Brendan O’Connor, *Racial Disparity in Natural Language Processing: A Case Study of Social Media African-American English* (June 30, 2017) (unpublished manuscript), <https://arxiv.org/pdf/1707.00061> [<https://perma.cc/DY5K-Z2ZP>]; Allison Koenecke et al., *Racial Disparities in Automated Speech Recognition*, 117 PROC. NAT’L ACAD. SCI. 7684, 7685 (2020); Holdsworth, *supra* note 80; BOMMASANI ET AL., *supra* note 51, at 24–25.

200. See generally Gabriel Nicholas & Aliya Bhatia, *Lost in Translation: Large Language Models in Non-English Content Analysis*, CTR. FOR DEMOCRACY & TECH. (May 2023), <https://cdt.org/insights/lost-in-translation-large-language-models-in-non-english-content-analysis> [<https://perma.cc/63YK-gCTJ>] (acknowledging that multilingual language models can expand economic opportunities and improve the web for all, but may also engage in erroneous content moderation, misinterpret employment and visa applications, and further entrench Anglocentrism).

201. OPENAI, *GPT-4 System Card*, at 21, n.27 (Mar. 23, 2023), <http://www.datascienceassn.org/sites/default/files/GPT-4%20System%20Card.pdf> [<https://perma.cc/PC8T-2T5J>].

202. Dave, *supra* note 188 (quoting an OpenAI staff member in a company forum).

For example, an AI program to verify signature matches that is trained on a dataset of signatures from a relatively homogenous state may have higher error rates when used to verify the authenticity of signatures on mail-in ballots in a relatively diverse state with large shares of Asian and Spanish surnames.²⁰³ Voice-assistance technology used to gather and process political information and to actually cast a ballot at the polls may be more accurate in standard varieties of English than in other languages or nonstandard dialects of English.²⁰⁴ A large share of Spanish misinformation stems from errors of automated systems in translating English news to Spanish.²⁰⁵ Facebook uses AI-powered content moderation tools, and one study showed that the company failed to alert users of falsehoods with seventy percent of misinformation in Spanish, compared with failing to add warning labels only twenty-nine percent of misinformation in English.²⁰⁶ California, which is required by the Voting Rights Act to provide election information in Spanish, has a chatbot that cannot provide some answers in Spanish that it provides in English.²⁰⁷

Large language models may help expand natural language processing capacity to additional languages due to the models' capacity to learn and adapt linguistic information.²⁰⁸ AI translation systems, however, "are good at translating other languages into English, but they struggle with rewriting English into other languages—especially those, like Korean, with non-Latin scripts."²⁰⁹

These realities can impair democratic engagement and participation within growing communities of color and can entrench elected officials and policy outcomes that more accurately reflect the interests of past democratic majorities. Both AI and democracy impact not just individuals but also groups.²¹⁰ Voting, for example, is relative—it is not exercised in isolation. If an AI-powered voice assistive technology in a polling place is designed in a way that makes it more helpful for non-Latino voters than for Latino voters, the voice assistive technology effectively dilutes the voting strength of Latino communities in electing their preferred candidates. Poorly designed assistive

203. See WEIDINGER ET AL., *supra* note 56, at 11; Brown et al., *supra* note 163, at 37–38.

204. SHEARER ET AL., *supra* note 183, at 3 ("Current NLP systems do not deal with linguistic variation. They are more accurate for standard varieties of a language than they are for non-standard varieties like African American Vernacular English.").

205. Memo from Laura MacCleery & Juan Espinoza of UnidosUS, to Author (Feb. 14, 2024) [hereinafter Memo from Laura MacCleery] (on file with the *Iowa Law Review*); see also Sanchez & Bennett, *supra* note 99.

206. *How Facebook Can Flatten the Curve of the Coronavirus Infodemic*, AVAAZ (Apr. 15, 2020), https://secure.avaaz.org/campaign/en/facebook_coronavirus_misinformation [<https://perma.cc/DCF5-T3T8>]; see Diana Enriquez, *The Algorithmic Divide: The Disparate Impact of Social Media News Curation on Spanish Speakers*, TECH POL'Y PRESS (Apr. 15, 2024), <https://www.techpolicy.press/the-algorithmic-divide-the-disparate-impact-of-social-media-news-curation-on-spanish-speakers> [<https://perma.cc/CYU6-MBCG>].

207. Memo from Laura MacCleery, *supra* note 205 (showing screenshot of chatbot error in Spanish from the State of California on January 24, 2024).

208. BOMMASANI ET AL., *supra* note 51, at 24.

209. Dave, *supra* note 188.

210. Jungherr, *supra* note 8, at 3 ("[A]t the *group* level, AI impacts equality of rights among different groups of people in society . . .").

technology that is calibrated (intentionally or unintentionally) in a way that benefits some racial groups at the polls has a dilutive effect on other racial groups because of racially polarized voting. The voice assistive technology harms not only the individual Latino who is unable to vote, but also Latinos who voted and are in coalition with those excluded.

3. Racially Biased Content Moderation

The architecture of AI can even facilitate bias within the AI-powered tools used to identify and limit racial disinformation and xenophobic content sometimes produced by AI.

Social media platforms often develop and deploy automated content moderation tools with algorithms that rely on large amounts of training data that are susceptible to racial bias.²¹¹ As a result, Black users experience disproportionate levels of account removals by social media platforms.²¹² Facebook employees, for example, revealed that internal company research showed that Black U.S. Instagram users were about fifty percent more likely “to have their accounts automatically disabled by the moderation system than those whose activity indicated they were white.”²¹³ Another study of Facebook, Instagram, Twitter (now known as X), YouTube, and other social media sites by University of Michigan researchers found that Black users were sixty percent more likely to have their accounts removed than white users.²¹⁴

Many platforms’ automated content moderation systems fail to recognize cultural nuances—thus disproportionately silencing users of color.²¹⁵ For example, some Black users claim that they “can’t talk about racism on Facebook without risking having their posts removed and being locked out of their accounts” and that Facebook has arbitrarily decided “that talking about racism is racist.”²¹⁶ TikTok apologized after its algorithm flagged as inappropriate

211. CAREY SHENKMAN, DHANARAJ THAKUR & EMMA LLANSÓ, CTR. FOR DEMOCRACY & TECH., *DO YOU SEE WHAT I SEE? CAPABILITIES AND LIMITS OF AUTOMATED MULTIMEDIA CONTENT ANALYSIS* 26–28 (2021).

212. See Oliver L. Haimson, Daniel Delmonaco, Peipei Nie & Andrea Wegner, *Disproportionate Removals and Differing Content Moderation Experiences for Conservative, Transgender, and Black Social Media Users: Marginalization and Moderation Gray Areas*, 2021 PROC. ACM ON HUM.-COMPUT. INTERACTION 466:1, 466:6.

213. Olivia Solon, *Facebook Ignored Racial Bias Research, Employees Say*, NBC NEWS (July 23, 2020, 2:29 PM), <https://www.nbcnews.com/tech/technews/facebook-management-ignored-internal-research-showing-racial-bias-current-former-n1234746> [<https://perma.cc/PWS3-6EKA>]; see also Daphne Keller, *Toward a Clearer Conversation About Platform Liability*, KNIGHT’S FIRST AMEND. INST. (Apr. 6, 2018), <https://knightcolumbia.org/content/toward-clearer-conversation-about-platform-liability> [<https://perma.cc/DF5Y-77CN>] (“[A] growing body of evidence suggests that they disproportionately harm vulnerable or disfavored groups.”).

214. See Haimson et al., *supra* note 212, at 466:1–4.

215. Brief for Law.’s Comm. for C.R. Under L. et al. as Amici Curiae Supporting Neither Party at 28–29, *Gonzalez v. Google LLC*, 598 U.S. 617 (2023) (No. 21-1333) (“The content moderation systems of many platforms lack the ability to recognize cultural nuances not rooted in a white, male, straight context, resulting in disproportionate silencing of users of color and other underserved groups.”).

216. Jessica Guynn, *Facebook While Black: Users Call It Getting ‘Zucked,’ Say Talking About Racism Is Censored as Hate Speech*, USA TODAY (Apr. 24, 2019, 7:26 AM), <https://www.usatoday.com/story>

phrases such as “Black Lives Matter,” “supporting black excellence,” and “pro black” but did not flag “white supremacy,” “supporting white excellence,” and “pro white.”²¹⁷ One study found that Twitter content moderation algorithms were more than twice as likely to flag as offensive tweets written in Black vernacular.²¹⁸

The racial harms discussed above of AI being used to impersonate people of color are enhanced by the racial biases of AI-powered video deepfake detection systems, which are often trained on an insufficiently robust number of images of people of color.²¹⁹ One study of three popular deepfake detectors found up to a 10.7% difference in error rate depending on gender and race.²²⁰ “[I]n a real-world scenario, facial profiles of female Asian or female African are 1.5-3 times more likely to be mistakenly labeled as fake than profiles of the male Caucasian. For large scale commercial applications, this would indicate bias against millions of people.”²²¹ The detectors had the highest error rates on videos with darker Black faces, particularly Black males.²²²

C. THE CHILLING EFFECTS OF AI SURVEILLANCE IN OVERPOLICED AND IMMIGRANT COMMUNITIES

Racial harms to democracy also arise from warrantless and often unregulated government surveillance that chills the political speech of people of color and

/news/2019/04/24/facebook-while-black-zucked-users-say-they-get-blocked-racism-discussion/2859593002 [https://perma.cc/GRG9-A6XZ]; see also Tracy Jan & Elizabeth Dvoskin, *A White Man Called Her Kids the N-Word. Facebook Stopped Her from Sharing It*, WASH. POST (July 31, 2017, 6:02 PM), https://www.washingtonpost.com/business/economy/for-facebook-erasing-hate-speech-proves-a-daunting-challenge/2017/07/31/922d9bc6-6e3b-11e7-9c15-177740635e83_story.html (on file with the *Iowa Law Review*) (discussing one woman’s practical issues with the site).

217. Charlotte Colombo, *TikTok Has Apologized for a ‘Significant Error’ After a Video That Suggested Racial Bias in Its Algorithm Went Viral*, BUS. INSIDER (July 8, 2021, 12:28 PM), <https://www.businessinsider.com/tiktok-racism-algorithm-apology-creator-marketplace-ziggy-tyler-2021-7> [https://perma.cc/C472-MX22].

218. Maarten Sap et al., *The Risk of Racial Bias in Hate Speech Detection*, 57 PROC. ANN. MEETING ASS’N FOR COMPUTATIONAL LINGUISTICS 1668, 1671 (2019); see also Thomas Davidson, Debasmita Bhattacharya & Ingmar Weber, *Racial Bias in Hate Speech and Abusive Language Detection Datasets*, 2019 PROC. 3D WORKSHOP ON ABUSIVE LANGUAGE ONLINE 25, 31 (“[W]e found that tweets in the black-aligned dataset were significantly more likely to be classified as harassment.”).

219. Jim Nash, *Bias in Facial Recognition Is Handicapping Deepfake Detection*, BIOMETRIC UPDATE (May 17, 2021, 7:01 PM), <https://www.biometricupdate.com/202105/bias-in-facial-recognition-is-handicapping-deepfake-detection> [https://perma.cc/H7VY-3RG9] (“[D]eepfake detectors are less able to spot fraudulent images and video of people of color.”).

220. Loc Trinh & Yan Liu, *An Examination of Fairness of AI Models for Deepfake Detection*, 13 PROC. INT’L JOINT CONF. ON A.I. 567, 568 (2021) (“Using facial datasets balanced by gender and race, we find that classifiers designed to detect deepfakes have large predictive disparities across racial groups, with up to 10.7% difference in error rate.”); see also Patrick Hall & Andrew Burt, Presentation at 4th Workshop on Payments, Lending and Innovations in Consumer Finance, Do Deepfakes Discriminate? Auditing a Deepfake Detection System for Systemic Bias, at slide 10 (Oct. 27, 2022), available at <https://www.philadelphiafed.org/-/media/frbp/assets/events/2022/consumer-finance/hall-deep-fakes-presentation-102722.pdf> [https://perma.cc/EB7T-K7Z6] (showing that for every 1000 deepfakes detected with white faces, one expects 694 deepfakes with South Asian faces and 821 deepfakes with Black faces to be detected).

221. Trinh & Liu, *supra* note 220, at 4–5.

222. *Id.* at 3.

their allies.²²³ “Location tracking, the related use of facial-recognition tools, and targeted surveillance of groups and protestors exercising their fundamental rights and freedoms are paramount data-privacy practices disproportionately impacting African Americans” and other communities of color,²²⁴ and are facilitated by various forms of AI-powered tools.

For example, in June 2020—a few weeks after white Minneapolis police officer Derek Chauvin killed George Floyd—private company Mobilewalla released an eighteen-slide deck entitled “George Floyd Protestor Demographics: Insights Across [Four] Major US Cities.”²²⁵ Mobilewalla obtained data from 16,902 mobile phones at protests in Atlanta, Los Angeles, Minneapolis, and New York and the company used artificial intelligence and “a stew of location data, device IDs, and browser histories to predict a person’s demographics—including race, age, gender, zip code, or personal interests.”²²⁶ Mobilewalla’s CEO indicated that his company has access to data from three hundred million devices in the United States and location data for thirty to sixty percent of people located in the United States.²²⁷

Although Mobilewalla released the report to the public, the company and other vendors routinely obtain data from advertisers, internet service providers, and data brokers, analyze it, and sell it to companies and government entities (including law enforcement).²²⁸ The Department of Homeland Security, for example, has paid two companies millions of dollars to obtain cell phone location

223. See generally Vincent M. Southerland, *The Master’s Tools and a Mission: Using Community Control and Oversight Laws to Resist and Abolish Police Surveillance Technologies*, 70 UCLA L. REV. 2 (2023) (examining local laws that empower local officials to oversee how law enforcement deploys surveillance technologies); UNIDOS US, WRITTEN COMMENTS ON LAW ENFORCEMENT USE OF FACIAL RECOGNITION TECHNOLOGY, BIOMETRIC SURVEILLANCE TECHNOLOGIES, PREDICTIVE ALGORITHMS, AND DATA STORAGE AND ACCESS 2 (2024), <https://unidosus.org/publications/unidosus-comments-addressing-concerns-on-law-enforcement-technologies> [<https://perma.cc/QU29-2Z5K>] (“[Surveillance] can also be a source of systematic civil rights and liberties violations.”); *Ban Facial Recognition Interactive Map*, BAN FACIAL RECOGNITION, <https://www.banfacialrecognition.com/map> [<https://perma.cc/75KW-NEQB>] (map showing jurisdictions that use facial recognition as well as those that have adopted bans on the use of facial recognition); JUNEJA, *supra* note 8, at 37–38 (detailing discriminatory surveillance that chills participation by groups of individuals traditionally subject to discrimination).

224. Anita L. Allen, *Dismantling the “Black Opticon”: Privacy, Race Equity, and Online Data-Protection Reform*, 131 YALE L.J.F. 907, 921 (2022).

225. *George Floyd Protestor Demographics: Insights Across 4 Major US Cities*, MOBILEWALLA (June 2020), <https://f.hubspotusercontent40.net/hubfs/4309344/MW%20Protester%20Demographic%20Analysis.pdf> [<https://web.archive.org/web/20200626002332/https://f.hubspotusercontent40.net/hubfs/4309344/MW%20Protester%20Demographic%20Analysis.pdf>].

226. Caroline Haskins, *Almost 17,000 Protesters Had No Idea a Tech Company Was Tracing Their Location*, BUZZFEED NEWS (June 25, 2020, 1:40 PM), <https://www.buzzfeednews.com/article/carolinehaskins1/protests-tech-company-spying> [<https://perma.cc/EY3F-F644>].

227. *Id.*

228. *Id.* (“Mobilewalla sells aggregated versions of that stuff back to advertisers.”); *The New Invisible Hand? The Impact of Algorithms on Competition and Consumer Rights: Hearing Before the Subcomm. on Competition Pol’y, Antitr. & Consumer Rts. of the S. Comm. on the Judiciary*, 117th Cong. 47 (2023) [hereinafter *Hewitt, Congressional Testimony*] (statement of Damon Hewitt, President & Exec. Dir., Law.’s Comm. for C.R.) (“The same broker has told Congress that data it is [sic] has sold to other brokers has subsequently been sold to law enforcement and the military.”).

data without a warrant in order to track the movement of immigrants and U.S. citizens in the United States.²²⁹

AI analysis of social media also facilitates low-cost mass surveillance of protesters of color by state and local law enforcement with little oversight.²³⁰ Language models can be used to build text classification tools that use natural language processing and sentiment analysis to assess the tenor of posts and identify patterns that humans may not detect.²³¹ For example, following the killing of George Floyd, police used artificial intelligence tools that monitored social media posts to track demonstrators' latest whereabouts and actions.²³² Similarly, the company Geofeedia developed and promoted to law enforcement its Geofeed Streamer tool, which used data from Twitter, Facebook, and Instagram to monitor Black Lives Matter activists.²³³ College police on hundreds of campuses have also purchased and used AI-powered social media monitoring tools for surveilling, "mitigating," and "forestalling" student protests on various issues, including demonstrations to remove Confederate statues and recent pro-Palestinian protests.²³⁴ Law enforcement uses AI-powered facial recognition tools to try to identify protesters by comparing photos scraped from social media pages with images from drones and closed-circuit television systems.²³⁵

Individuals subject to this surveillance may reasonably presume that exercising their speech rights will result in harms and self-censor to avoid problems.²³⁶ The "automated, continuous, and unspecific collection, retention,

229. Shreya Tewari & Fikayo Walter-Johnson, *New Records Detail DHS Purchase and Use of Vast Quantities of Cell Phone Location Data*, ACLU (July 18, 2022), <https://www.aclu.org/news/privacy-technology/new-records-detail-dhs-purchase-and-use-of-vast-quantities-of-cell-phone-location-data> [<https://perma.cc/3NCG-UUW6>].

230. Adrian Shahbaz & Allie Funk, *Social Media Surveillance: Freedom on the Net 2019 Key Finding: Governments Harness Big Data for Social Media Surveillance*, FREEDOM HOUSE (2019), <https://freedomhouse.org/report/freedom-on-the-net/2019/the-crisis-of-social-media/social-media-surveillance> [<https://perma.cc/L6MU-PDF5>].

231. *Id.*; WEIDINGER ET AL., *supra* note 56, at 28 ("[Language models] can be used to build text classification tools that can . . . achieve high accuracy in . . . identifying, for example, political dissent at scale.").

232. Sam Biddle, *Police Surveilled George Floyd Protests with Help from Twitter-Affiliated Startup Dataminr*, INTERCEPT (July 9, 2020, 2:00 PM), <https://theintercept.com/2020/07/09/twitter-dataminr-police-spy-surveillance-black-lives-matter-protests> (on file with the *Iowa Law Review*).

233. In response to public backlash the social media platforms stopped sharing the data. *Hewitt, Congressional Testimony, supra* note 228, at 48 ("Facebook, Twitter, and Instagram provided user data to Geofeedia, a social media monitoring product that was marketed to law enforcement agencies to surveil civil rights activists.").

234. Ari Sen & Derëka K. Bennett, *Tracked: How Colleges Use AI to Monitor Student Protests*, DALL. MORNING NEWS (Sept. 20, 2022), <https://interactives.dallasnews.com/2022/social-sentinel> (on file with the *Iowa Law Review*); Chris Morris, *Why Facial Recognition Technology Makes These Campus Protests Different from Those in the Past*, FAST CO. (May 2, 2024), <https://www.fastcompany.com/91116791/facial-recognition-technology-campus-protests-police-surveillance-gaza> [<https://perma.cc/96UC-5KEE>].

235. Joy Buolamwini, *We Must Fight Face Surveillance to Protect Black Lives: An Urgent Letter from the Algorithmic Justice League*, MEDIUM (June 3, 2020), <https://onezero.medium.com/we-must-fight-face-surveillance-to-protect-black-lives-5ffcdob4c28a> [<https://perma.cc/J3P9-LJJC>].

236. Daniel J. Solove, *A Taxonomy of Privacy*, 154 U. PA. L. REV. 477, 488 (2006) (explaining that modern privacy problems such as digital surveillance by law enforcement may chill behavior,

and analysis of digital traces by state and corporate actors” that is “always traceable”²³⁷ may chill speech from Black Lives Matter activists and others from vulnerable communities.²³⁸ As Freedom House wrote:

The chilling effect on free expression caused by increased surveillance is well documented. Activists and journalists who might otherwise hold governments to account for wrongdoing are more inclined to self-censor, while dissidents and members of marginalized communities will think twice about discussing their political opinions online to avoid arrests or travel restrictions.²³⁹

Those who are not U.S. citizens may be particularly chilled in their political expression.²⁴⁰ In 2020, a graduate student indicated on Twitter that they had wanted to say something about the president’s executive order to suspend temporary work visa renewals but did not for fear of negatively affecting the student’s visa renewal.²⁴¹ Another study found that Muslims in America were substantially more chilled in engaging in online political activities than the general public after being told of government surveillance of their online activities.²⁴²

A system in which communities of color are chilled from political participation because they are “hyper visible and exposed to systems of surveillance”²⁴³ is not a well-functioning multiracial democracy that equally facilitates participation by all communities. AI applications become tools of racial control when they are misused in surveilling protesters of color.²⁴⁴

making people “less likely to attend political rallies or criticize popular views”); Jonathon W. Penney, *Internet Surveillance, Regulation, and Chilling Effects Online: A Comparative Case Study*, INTERNET POL’Y REV. 8 (May 26, 2017), <https://policyreview.info/articles/analysis/internet-surveillance-regulation-and-chilling-effects-online-comparative-case> [<https://perma.cc/8HK2-G2W8>] (describing survey results finding that government surveillance chilled speech, “with [sixty-two percent] of respondents ‘much less likely’ ([twenty-two percent]) or ‘somewhat less likely’ ([forty percent]) [to] ‘speak or write about certain topics online’ due to such online monitoring”).

237. Moritz Büchi, Noemi Festic & Michael Latzer, *The Chilling Effects of Digital Dataveillance: A Theoretical Model and an Empirical Research Agenda*, BIG DATA & SOC’Y 1 (Jan. 6, 2022), <https://journal.sagepub.com/doi/10.1177/20539517211065368> [<https://perma.cc/BR5N-SDA6>].

238. *Id.* at 7 (“[I]ndividuals potentially affected by such discrimination may justifiably form beliefs of negative outcomes, which mediate how the sense of dataveillance ultimately impacts digital communication.”).

239. Shahbaz & Funk, *supra* note 230.

240. Büchi et al., *supra* note 237, at 6 (“[N]oncitizens may be much more wary of researching or posting content critical of the government.”); *see also* Haskins, *supra* note 226.

241. Büchi et al., *supra* note 237, at 1.

242. Elizabeth Stoycheff, Juan Liu, Kai Xu & Kunto Wibowo, *Privacy and the Panopticon: Online Mass Surveillance’s Deterrence and Chilling Effects*, 21 NEW MEDIA & SOC’Y 602, 611–12 (2019).

243. *Conversations for Change*, *supra* note 78 (“Coded exposure examines the way that some technologies fail to see racial differences while others render racialized people hyper visible and exposed to systems of surveillance.”).

244. *See* BENJAMIN, *supra* note 78, at 139–40 (discussing the use of technological surveillance by the state as a tool of control of Black people).

D. RACIAL ENTRENCHMENT THROUGH ELECTION STRUCTURES

Artificial intelligence technologies pose significant risks of racial harm in the context of election structures. Without proper safeguards, AI models can automate racial bias in various election procedures. Local election offices that serve large populations of color are particularly vulnerable to AI-powered cyberattacks, nuisance open-record requests, mass frivolous voter challenges, and threats of violence against election workers.

1. Automating Bias in Election Administration

Election administrators can deploy AI tools for a wide range of seemingly positive functions. For example, some state and local election offices have adopted chatbots to answer voter questions.²⁴⁵ In the future, election administrators could deploy AI tools to decide where to allocate resources (e.g., location of polling sites, number of election workers and polling booths at particular sites), design ballots, translate election materials into various languages, automatically register eligible voters, monitor social media posts to detect election misinformation and polling place problems (e.g., broken machines, long lines, voter intimidation, vulnerability to violence), and much more.²⁴⁶ AI may also be used to conduct audits after elections to identify patterns and reveal why certain precincts had long lines or high rates of provisional ballot rejections.²⁴⁷

As discussed above, however, real risks arise from bias in foundation models and AI system design. California's chatbot that can provide some answers to election questions in English that it cannot provide in Spanish is but one example.²⁴⁸ To the extent that AI applications embed bias in other aspects of election administration, election officials risk expanding racial disparities in access to voting.

For several years, state and local election offices have used automated systems to identify and remove ineligible names from voter registration lists (e.g., voters who have died or moved) and to perform initial signature verifications for mail-in ballots.²⁴⁹ While there are debates about which tools used to perform these tasks mimic "human cognitive functions like problem-solving and learning,"²⁵⁰ or are simply "rudimentary data matching,"²⁵¹ most practices

245. Statement of Lawrence D. Norden, Executive Director, Elections & Gov't, Brennan Ctr. for Just. NYU Sch. of L., to U.S. Sen. Chuck Schumer et al. (Nov. 8, 2023) [hereinafter Statement of Lawrence D. Norden], <https://www.brennancenter.org/ourwork/research-reports/brennancenter-statement-fifth-bipartisan-senate-forum-artificial> [<https://perma.cc/VE5X-VUJ4>].

246. *Id.*; JUNEJA, *supra* note 8, at 17–20 (detailing electoral management body use of AI for polling place location and resource allocation determinations and other functions).

247. JUNEJA, *supra* note 8, at 31–32 (detailing the use of AI in post-electoral audits).

248. Memo from Laura MacCleery, *supra* note 205 (showing screenshot of chatbot error in Spanish from the State of California on January 24, 2024).

249. See Statement of Lawrence D. Norden, *supra* note 245, at 2.

250. *AI vs. Machine Learning*, *supra* note 48.

251. Edgardo Cortés, Lawrence Norden, Heather Frase & Mia Hoffman, *Safeguards for Using Artificial Intelligence in Election Administration*, BRENNAN CTR. FOR JUST. (Dec. 12, 2023), <https://www.brennancenter.org/ourwork/research-reports/safeguards-for-using-artificial-intelligence-in-election-administration>.

generally fall within the broad legal definition of AI adopted by the OECD and many other jurisdictions.²⁵²

Regardless of the specific technology deployed, automated purging systems have the potential to “perpetuate systemic biases that disparately impact communities of color.”²⁵³ One study of Wisconsin voters revealed that the rate at which voters were erroneously flagged as having moved was 141% higher for people of color than for whites.²⁵⁴ Name-matching algorithms have been found to render more mismatches for the most common names of Asian Americans than for the most common names of white Americans and Black Americans,²⁵⁵ partly because many Asian American names are short and often use the same vowels.²⁵⁶

Bias in automated verification of signatures on mail-in ballots poses similar challenges. In the 2020 presidential elections, forty-three percent of votes were cast via mail-in or absentee ballot.²⁵⁷ Before most of these ballots were counted, they were evaluated to ensure that the signature on the ballot matched the signature on file for the voter.²⁵⁸ A nonmatching signature was by far the

w.brennancenter.org/our-work/research-reports/safeguards-using-artificial-intelligence-election-administration [https://perma.cc/Z88g-W4FL] (“The Electronic Registration Information Center (ERIC), a multistate voter list maintenance effort, is one example of non-generative AI use in election administration. ERIC’s software employs AI to support voter roll management by searching for duplicate entries across many data sets. . . . ERIC’s ability . . . is considerably more advanced than earlier systems, such as the Interstate Voter Registration Crosscheck program, which utilized rudimentary data matching with limited date fields and led to high numbers of false positive identifications.”).

252. For a legal definition of AI, see *supra* Part II.

253. Statement of Jocelyn Benson, Mich. Sec’y of State, Submitted for the AI Insight Forum Focused on AI and Its Impact on Elections and Democracy, to U.S. Sen. Chuck Schumer et al. (Nov. 8, 2023) [hereinafter Statement of Jocelyn Benson], <https://www.schumer.senate.gov/imo/media/doc/Jocelyn%20Benson%20-%20Statement.pdf> [https://perma.cc/X8K5-QXDW].

254. Gregory A. Huber, Marc Meredith, Michael Morse & Katie Steele, Voter List Maintenance Errors and Their Racial Burden: Evidence from Wisconsin’s Supplemental Movers Poll Books 3, 16–17 (Dec. 23, 2020) (unpublished manuscript), <https://www.sas.upenn.edu/~marcmere/workingpapers/WIListMaintenance.pdf> [https://perma.cc/TPX9-LRLN] (finding that 9% of voters who cast ballots in Wisconsin in 2018 voted from an address flagged by officials as having moved, and that the false mover error rate of this group was 2.7% for whites and 6.5% for people of color).

255. Bender, *supra* note 9, at 505 (citing Alexandros Karakasidis & Evaggelia Pitoura, *Identifying Bias in Name Matching Tasks*, 22 INT’L CONF. ON EXTENDING DATABASE TECH. 626, 626 (2019)).

256. Karakasidis & Pitoura, *supra* note 255, at 628. Of the 162,253 distinct names in the U.S. Census Bureau’s database, the study identified those names for which at least ninety percent of those with the name belonged to a single racial group, and from those groups selected the fifty most common names from each racial group to test. *Id.* at 627. The study found negative bias rates for Asian Americans, relatively average error rates for whites, and a positive bias for the Black American names. *Id.* at 628.

257. Zachary Scherer, *What Methods Did People Use to Vote in the 2020 Election?: Majority of Voters Used Nontraditional Methods to Cast Ballots in 2020*, U.S. CENSUS (Apr. 29, 2021), <https://www.census.gov/library/stories/2021/04/what-methods-did-people-use-to-vote-in-2020-election.html> [https://perma.cc/VU66-75UG].

258. Bender, *supra* note 9, at 507.

most common reason for rejecting mail-in ballots, accounting for 32.8% of the 560,826 ballots rejected by election officials in 2020.²⁵⁹

Increasingly, initial signature matching is done by automated signature verification software.²⁶⁰ Algorithms evaluate the width, height, symmetry, and stroke directions of the mail-in ballot signature compared to those in the voter files,²⁶¹ and ballots that clear a certain “confidence threshold” are verified and counted in many counties, while those that fail to do so are flagged as possibly erroneous and are generally reviewed by humans.²⁶² The accuracy rate for signature verification software varies from as low as seventy-four percent up to about ninety-six percent.²⁶³

Voters of color experience higher rejection rates of mail-in ballots due to election officials’ decisions about nonmatching signatures.²⁶⁴ In Washington State, for example—where one county uses automated signature verification—a statewide audit found that mail-in ballots during the 2020 election were discarded due to problematic signatures four times more often for Black voters than for white voters.²⁶⁵

259. U.S. ELECTION ASSISTANCE COMM’N, ELECTION ADMINISTRATION AND VOTING SURVEY 2020 COMPREHENSIVE REPORT: A REPORT FROM THE U.S. ELECTION ASSISTANCE COMMISSION TO THE 117TH CONGRESS 14, 36 (2021).

260. Pares Dave & Andy Sullivan, *Factbox: U.S. Counties Using Automated Signature Verification Software*, REUTERS (Sept. 24, 2020), <https://www.reuters.com/article/idUSKCN26F1U4> (on file with the *Iowa Law Review*); Bender, *supra* note 9, at 508.

261. Sabri Ben-Achour, *Robots Will Be Verifying Some of Our Ballots. Can We Trust Them?*, MARKETPLACE: MARKETPLACE TECH, at 02:24 (Oct. 30, 2020), <https://www.marketplace.org/shows/marketplace-tech/vote-by-mail-ballots-mismatched-signatures-verification-software-disenfranchisement> [<https://perma.cc/KX3S-VWMS>]; Kyle Wiggers, *Automatic Signature Verification Software Threatens to Disenfranchise U.S. Voters*, VENTUREBEAT (Oct. 25, 2020, 10:25 AM), <https://venturebeat.com/2020/10/25/automatic-signature-verification-software-threatens-to-disenfranchise-u-s-voters> [<https://web.archive.org/web/20241009053353/https://venturebeat.com/ai/automatic-signature-verification-software-threatens-to-disenfranchise-u-s-voters>]; Bender, *supra* note 9, at 508; ROXANA ARJON ET AL., SIGNATURE VERIFICATION AND MAIL BALLOTS: GUARANTEEING ACCESS WHILE PRESERVING INTEGRITY 26–29 (2020), <https://law.stanford.edu/publications/signature-verification-and-mail-ballots-guaranteeing-access-while-preserving-integrity> [<https://perma.cc/HHC9-3NQC>].

262. Bender, *supra* note 9, at 508–09 (“In many jurisdictions, ballots flagged as a possible mismatch are manually examined by staff, while those the AI approves are not. Other counties require staff to verify each ballot, regardless of the AI’s results.”); ARJON ET AL., *supra* note 261, at 25–26.

263. Wiggers, *supra* note 261; Raashid Hussain, Ahsen Raza, Imran Siddiqi, Khurram Khurshid & Chawki Djeddi, *A Comprehensive Survey of Handwritten Document Benchmarks: Structure, Usage and Evaluation*, EURASIP J. ON IMAGE & VID. PROCESSING 18 (Dec. 24, 2015), <https://jivp-urasipjournalns.springeropen.com/articles/10.1186/s13640-015-0102-5> [<https://perma.cc/3ZMZ-EUZF>] (finding accuracy rates ranging from 73.4% for an algorithm trained on a one thousand-writer database to 96.7% for an algorithm trained on a 657-writer database).

264. Bender, *supra* note 9, at 509; Maya Lau & Laura J. Nelson, ‘Ripe for Error’: Ballot Signature Verification Is Flawed—and a Big Factor in the Election, L.A. TIMES (Oct. 28, 2020), <https://www.latimes.com/california/story/2020-10-28/2020-election-voter-signature-verification> (on file with the *Iowa Law Review*).

265. Mike Baker, *Rejected Mail Ballots Are Showing Racial Disparities*, N.Y. TIMES (Feb. 2, 2022), <https://www.nytimes.com/2022/02/02/us/mail-voting-black-latino.html> (on file with the *Iowa Law Review*). The audit also found that Native American, Latino, and Asian and Pacific Islander voters were more likely to have their ballots discarded due to signature issues than whites. See OFF.

With automated systems, racial disparities in the rate of nonmatching signatures could stem from various sources. Signature-matching tools are generally trained on English handwriting, and thus voters who write primarily in other languages “may be at greater risk of having their ballot rejected.”²⁶⁶ In training models, vendors often use training data that differs from the actual constituents’ data in a particular jurisdiction, and thus the model may work less effectively on it.²⁶⁷

Almost all jurisdictions that use automated tools to review registration rolls and/or mail ballot signatures have a human review the decision of the automated system.²⁶⁸ The bias of automated systems is also relative—automated systems may exhibit less bias than humans in reviewing a large number of items. That said—bias is bias—and presents a problem if votes cast from communities of color are subject to additional scrutiny, or if racial bias stems from other automated election administration functions.

2. Racially Targeted Attacks on Local Election Offices

In addition to introducing inadvertent racial bias in election administration, AI could make it easier for foreign adversaries, domestic political operatives, and individual trolls to try to shift election outcomes through targeted attacks on county election offices that serve large communities of color.

In the United States, elections are administered on the local level, and in about four hundred counties people of color are now a majority.²⁶⁹ An AI attack on a handful of county election offices that serve metro areas in closely contested swing states with large populations of color (e.g., Orlando, Atlanta, Detroit, Las Vegas, Charlotte, Philadelphia, Norfolk, Milwaukee) could divert resources, increase vote wait times,²⁷⁰ make voting more difficult, and potentially shift the outcome of a statewide or even presidential election. In

OF THE WASH. STATE AUDITOR, EVALUATING WASHINGTON’S BALLOT REJECTION RATES 19 (2022), https://sao.wa.gov/sites/default/files/audit_reports/PA_Evaluating_WA_Ballot_Rejections_ar-10_29711.pdf [<https://perma.cc/9NHN-9S6U>].

²⁶⁶ Bender, *supra* note 9, at 509; *see also* Wiggers, *supra* note 261; Sabri Ben-Achour, *supra* note 261.

²⁶⁷ Cortés et al., *supra* note 251.

²⁶⁸ Dave & Sullivan, *supra* note 260.

²⁶⁹ Stef W. Kight & Connor Rothschild, *More Than 400 U.S. Counties Are Now Minority White*, AXIOS (Aug. 15, 2021), <https://www.axios.com/2021/08/15/diversity-majority-minority-white-american-census> (on file with the *Iowa Law Review*); Jens Manuel Krogstad, *Reflecting a Demographic Shift, 109 U.S. Counties Have Become Majority Nonwhite Since 2000*, PEW RSCH. CTR. (Aug. 21, 2019), <https://www.pewresearch.org/short-reads/2019/08/21/u-s-counties-majority-nonwhite> [<https://perma.cc/DXX2-EAE6>]; *Counties with Majority BIPOC Population*, REDISTRICTING DATA HUB, <https://redistrictingdatahub.org/tools/support/local-redistricting/counties-with-majority-bipo-c> [<https://perma.cc/KFH2-XK5F>] (listing 391 counties in the United States with a majority Black, Indigenous, and people of color population).

²⁷⁰ Voting wait times are already longer for Black and Latino voters. *See, e.g.*, Hanna Klain, Kevin Morris, Rebecca Ayala & Max Feldman, *Waiting to Vote*, BRENNAN CTR. FOR JUST. (June 3, 2020), <https://www.brennancenter.org/our-work/research-reports/waiting-vote> [<https://perma.cc/U3VM-4JDM>] (reporting that in the 2018 general election, “Latino voters waited on average [forty-six] percent longer than white voters, and Black voters waited on average [forty-five] percent longer than white voters”).

the alternative, an attack on a handful of election offices in majority nonwhite rural counties—which may have fewer resources to defend against AI-powered attacks—could suppress enough votes to change election outcomes in closely contested states like Arizona, North Carolina, and Georgia.

For example, in past elections, some political operatives have targeted mass voter challenges at counties with higher-than-average populations of color.²⁷¹ Moving forward, political operatives could use AI to scrape substandard data from the web about voters in counties with large populations of color, engage in incomplete data matching against voter registration records, automatically complete voter registration challenge forms, and overwhelm county election officials with mass voter challenges based on unreliable evidence.²⁷²

Since 2020, election fraud conspiracy adherents have flooded election officials with open-record requests, which have required that election officials divert resources away from other priorities just before Election Day.²⁷³ Rather than manually crafting an open-record request and copying and pasting identical language onto multiple forms, AI-powered synthetic text generators will allow a single troll to file thousands of unique requests on various topics to a single county office, each of which seems to come from a different person.²⁷⁴

Similarly, AI can be used to target harassment and intimidation at election workers in jurisdictions that serve large communities of color. Such harassment has been trending upward generally,²⁷⁵ and disaggregated data from one survey show that election workers are targeted at higher rates when they serve in majority people of color jurisdictions.²⁷⁶

271. JONATHAN BRATER, *VOTER PURGES: THE RISKS IN 2018*, at 6 (2018), <https://www.brennancenter.org/our-work/research-reports/voter-purges-risks-2018> [<https://perma.cc/5DAP-5R5Y>]; *id.* at 11 n.57 (“[I]n several states a large percentage of the counties [Public Interest Legal Foundation] targeted had higher minority populations than the state average, including Alabama (12/12), Louisiana (2/2), Mississippi (17/19) and New York (4/6).”).

272. See Statement of Lawrence D. Norden, *supra* note 245, at 4; see also Jane C. Timm, *Inside the Right's Effort to Build a Voter Fraud Hunting Tool*, NBC NEWS (Aug. 17, 2023), <https://www.nbcnews.com/politics/2024-election/conservatives-voter-fraud-hunting-tool-eagleai-cleta-mitchell-rcna97327> [<https://perma.cc/SX96-ZFEg>] (“Activists are currently testing a computer program called EagleAI NETWORK, a database loaded with voter rolls and other records that promises to quickly churn through the data and find registrations that may be suspect based on other sources.”); Michael Waldman, *EagleAI Isn't So Intelligent*, BRENNAN CTR. FOR JUST. (Sept. 8, 2023), <https://www.brennancenter.org/our-work/analysis-opinion/eagleai-isnt-so-intelligent> [<https://perma.cc/YJ7-5JWM>].

273. Lawrence Norden & Gowri Ramachandran, *Artificial Intelligence and Election Security*, BRENNAN CTR. FOR JUST. (Oct. 5, 2023), <https://www.brennancenter.org/our-work/research-reports/artificial-intelligence-and-election-security> [<https://perma.cc/ESQ3-WZLK>].

274. *Id.*

275. Chelsey Cox, *Threats Against Election Workers Driving Exodus of Veteran Staff, Experts Say*, CNBC (Nov. 7, 2023, 5:10 PM), <https://www.cnn.com/2023/11/07/election-worker-threats-drive-exodus-from-profession.html> [<https://perma.cc/EE33-7ND3>].

276. Ruby Edlin & Lawrence Norden, *Election Officials in Communities of Color Face More Abuse*, BRENNAN CTR. FOR JUST. (July 17, 2024), <https://www.brennancenter.org/our-work/analysis-opinion/election-officials-color-face-more-abuse> [<https://perma.cc/F5GT-WWKR>] (finding that election officials that serve in majority people of color counties were “twice as likely to have been threatened [as] their counterparts who serve in majority-white counties”); see also Ishena Robinson, *Black Poll Workers Need Protecting to Keep America's Democracy Working*, NAACP LDF (Jan.

The lower-tech attacks deployed against Georgia election workers Ruby Freeman and Wandrea ArShaye “Shaye” Moss—both Black—provide a roadmap of possible tactics. Rudy Giuliani narrated snippets of a “smoking gun” surveillance video accusing Freeman of engaging in fraud in handling ballots during the 2020 presidential election (state and federal investigators quickly debunked Giuliani’s claims).²⁷⁷ Later, a fake Instagram account was erected purporting to be from Freeman in which she appeared to confess to election fraud.²⁷⁸ As a result, she experienced:

[A] torrent of threats, accusations and racism; messages from people who said she should be hanged for treason, or lynched; people who fantasized about hearing the sound of her neck snap. . . . The harassment got so bad that the F.B.I. told Ms. Freeman she was not safe in the home where she had lived for years.²⁷⁹

In the future, deepfake video, synthetic social media content appearing to come from thousands of people, and other AI applications could be targeted to attack the credibility of and threaten election workers like Ruby Freeman, and deter them from serving as election workers.²⁸⁰

Election offices in communities of color are also vulnerable to cyberattacks.²⁸¹ AI allows hackers (including those without deep technical expertise) to generate code designed to evade detection, analyze large datasets, and deploy a cyberattack that more easily overcomes cyber defenses.²⁸² Deepfake

5, 2023), <https://www.naacpldf.org/black-poll-workers-protect-democracy> [<https://perma.cc/YEB2-3BUU>]; GOWRI RAMACHANDRAN ET AL., BRENNAN CTR. FOR JUST., INTIMIDATION OF STATE AND LOCAL OFFICE HOLDERS 10 (2024), <https://www.brennancenter.org/our-work/research-reports/intimidation-state-and-local-officeholders> [<https://perma.cc/SE3V-M7EM>] (“Twenty-five percent of [local] officeholders of color reported threats, compared with [eighteen] percent of white officeholders.”); *id.* at 8 (showing that fifty-five percent of state legislators of color reported being subjected to hostile comments about race or ethnicity, compared to only seventeen percent of white state legislators).

277. David Wickert, ‘Smoking Gun’ Video of Georgia Vote Count Is Now Evidence Against Trump, *AJC POL.* (Aug. 5, 2023), <https://www.ajc.com/politics/smoking-gun-video-now-evidence-against-trump/J6ORVROLMRBPZHK2DYALIZJ624> [<https://perma.cc/6BXF-5FXK>].

278. Sam Levine, *Two Georgia Election Workers Cleared of Wrongdoing in 2020 Elections*, *GUARDIAN* (June 23, 2023), <https://www.theguardian.com/us-news/2023/jun/23/georgia-election-workers-cleared-trump-giuliani-vote-2020> [<https://perma.cc/RB2D-AABC>].

279. Eileen Sullivan, *Election Worker Tells Jury: ‘Giuliani Just Messed Me Up’*, *N.Y. TIMES* (Dec. 13, 2023), <https://www.nytimes.com/2023/12/13/us/politics/rudy-giuliani-ruby-freeman.html> (on file with the *Iowa Law Review*).

280. See, e.g., Carrie Levine, *Poll Worker Shortage Could Suppress Black Vote in November*, *CTR. FOR PUB. INTEGRITY* (Sept. 4, 2020), <https://publicintegrity.org/inside-publici/newsletters/poll-worker-shortage-could-suppress-black-vote-in-november> [<https://perma.cc/5ES8-ZC6P>]; see also Lucien Bruggeman, *Judge Orders Immediate Enforcement of Georgia Election Workers’ \$148M Judgment Against Giuliani*, *ABCNEWS* (Dec. 20, 2023), <https://abcnews.go.com/Politics/judge-agrees-expedite-georgia-election-workers-148m-judgment/story?id=105824432> [<https://perma.cc/L4CB-F85T>].

281. See JUNEJA, *supra* note 8, at 45–46 (discussing the harms of AI-enabled higher quality phishing attempts and information flooding).

282. Manheim & Kaplan, *supra* note 8, at 136; Bender, *supra* note 9, at 515 (“AI can also help hackers overcome barriers to widespread election hacking.”); Statement of Lawrence D. Norden, *supra* note 245, at 3 (“AI will also change how software is engineered. It can make cyberattacks

voicemails purporting to come from election supervisors or system vendors could facilitate phishing operations that entice election workers to provide login credentials or to open links to malware that alters or disables the voter registration database, recording system, or tabulating system.²⁸³ Even if no votes are altered, such attacks could slow down the voting process, lower voter of color turnout, and shift outcomes in a closely contested election.

IV. PRINCIPLES TO OVERCOME AI HARMS TO DEMOCRACY

As detailed above, AI technologies pose real challenges to racially inclusive democracy. Not only will some entrenched politicians and their supporters use AI applications to contain the political influence of growing communities of color, but due to training and design issues AI models could inadvertently disadvantage voters of color and fortify dominant culture and policy preferences of the past.

Unfortunately, the Voting Rights Act and other laws are unlikely to fully address these challenges. The Voting Rights Act prohibits voter intimidation,²⁸⁴ for example, and might be invoked to prevent deepfakes that use threats to discourage voters from casting a ballot (e.g., “you risk losing your Medicaid benefits if you vote by mail”), but claims based on deception alone may be less effective (e.g., “voting is on Wednesday, not Tuesday”).²⁸⁵ While the Federal Communications Commission recently outlawed robocalls with AI-generated voices to mobile phones,²⁸⁶ the Federal Election Commission has not yet acted on opportunities to regulate AI deepfake video and audio used to misrepresent federal candidates, and just over a third of states have done so.²⁸⁷ Some states

bigger, quicker, and sneakier to outsmart existing software security measures.”); WEIDINGER ET AL., *supra* note 56, at 27–28.

283. See Brandt, *supra* note 114 (“Chatbots, deepfake audio, and persuasive generated text could also enable phishing operations that are more personalized and convincing.”); ELECTION ASSISTANCE COMM’N, AI TOOLKIT FOR ELECTION OFFICIALS 6 (2023) (“Cybersecurity Risks: During its testing of GPT-4, OpenAI found that ChatGPT was effective at drafting targeted phishing emails.”); see also Statement of Jocelyn Benson, *supra* note 253.

284. Voting Rights Act of 1965, Pub. L. No. 89-110, § 11(b), 79 Stat. 437, 443 (codified at 42 U.S.C. § 1973i(b)).

285. See, e.g., Nat’l Coal. Black Civic Participation v. Wohl, 661 F. Supp. 3d 78, 92–93 (S.D.N.Y. 2023) (finding that robocalls falsely stating that voting-by-mail would result in a voters’ information being used by police and debt collectors were intimidating and thus violated Voting Rights Act § 11(b)).

286. Press Release, FCC, FCC Makes AI-Generated Voices in Robocalls Illegal (Feb. 8, 2024), <https://docs.fcc.gov/public/attachments/DOC-400393A1.pdf> [<https://perma.cc/E2TA-G7VH>]; see also Panditharatne, *supra* note 75 (“Federal law also continues to allow political robocalls to landlines — including those that use voice-cloning AI — without prior consent. (Landlines are disproportionately used by Americans over the age of 65.)”).

287. See *Artificial Intelligence (AI) in Elections and Campaigns*, NAT’L CONF. STATE LEGISLATURES (July 15, 2024), <https://www.ncsl.org/elections-and-campaigns/artificial-intelligence-ai-in-electons-and-campaigns> [<https://perma.cc/RX42-RPHH>] (listing nineteen state statutes governing the use of AI in elections and campaigns, and listing state legislation introduced in 2023 and 2024 that would do the same). See, e.g., S.B. 751, 86th Leg. (Tex. 2019); Elections: deceptive audio or visual media, A.B. 730, 2019–2020 Leg., Reg. Sess. (Cal. 2019); Electioneering Communications—Use of Synthetic Media, S.B. 5152, 68th Leg., Reg. Sess. (Wash. 2023); H.F. 1370, 93d Leg. (Minn. 2023); H.B. 5144, 2023–2024 Sess. (Mich. 2023).

have traditional laws prohibiting deceptive practices in elections, but only a few are sufficiently broad to apply to online voting deception.²⁸⁸

Depending on the fact pattern, a deepfake victim may be able to bring a legal claim for defamation, false light, right of publicity, or intentional infliction of emotional distress,²⁸⁹ but would face significant challenges (e.g., difficulty in identifying perpetrators, immunity claims by tech platforms under Section 230 of the Communications Decency Act,²⁹⁰ hurdles to collecting on a judgment).²⁹¹ Even if successful, an individual lawsuit would not fully address larger harms to democracy such as racial impersonation and infiltration, targeted harassment of election workers in jurisdictions of color, depressed voter turnout among communities of color, and increased racial polarization and cultural anxiety.

Most jurisdictions lack laws that regulate local law enforcement's warrantless surveillance using AI-powered analytics of mobile phone location data or social media content.²⁹² While the European Union regulates data privacy²⁹³ and will soon ban the use of AI for "cognitive [behavioral] manipulation of people or specific vulnerable groups,"²⁹⁴ the United States lacks such laws.²⁹⁵

The failure is not simply that of legislative branches, but also the U.S. Supreme Court. For example, in *Brnovich v. Democratic National Committee*, Justice Alito created five factors to govern a court's analysis of Section 2 of the Voting Rights Act, one of which was the "degree to which a voting rule departs from what was standard practice when [Section 2] was amended in 1982."²⁹⁶ Election practices in 1982, however, do not serve as an appropriate baseline

288. See Gilda R. Daniels, *Voter Deception*, 43 IND. L. REV. 343, 370–71 (2010).

289. See Chesney & Citron, *supra* note 9, at 1793–802.

290. 47 U.S.C. § 230; see also Spencer Overton & Catherine Powell, *The Implications of Section 230 for Black Communities*, 66 WM. & MARY L. REV. 1, 25–30 (2024).

291. See Overton & Powell, *supra* note 290, at 5.

292. See generally Southerland, *supra* note 223 (outlining the use of and effort to abate new surveillance technologies by police). In recent years, the F.B.I. created guidelines on warrantless surveillance, but has since violated them. See Charlie Savage, *F.B.I. Violated Surveillance Program Rules After George Floyd Protests and Jan. 6 Attack*, N.Y. TIMES (May 19, 2023), <https://www.nytimes.com/2023/05/19/us/politics/fbi-violated-surveillance-program-rules.html> (on file with the *Iowa Law Review*).

293. See generally Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119/1).

294. *EU AI Act: First Regulation on Artificial Intelligence*, EUR. PARLIAMENT (June 18, 2024, 4:29 PM), <https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence> [<https://perma.cc/6FBR-RBZV>]. As of January 2024, the final act was in the process of being finalized and approved by member states. *Germany to Approve EU's Planned AI Act*, REUTERS (Jan. 30, 2024, 2:52 PM), <https://www.reuters.com/technology/germany-will-approve-eus-planned-ai-act-sources-2024-01-30> (on file with the *Iowa Law Review*).

295. See Allen, *supra* note 224, at 908 ("Existing civil-rights laws and doctrines are not yet applied on a consistent basis to combat the serious discrimination and inequality compounded by the digital economy. Existing common law, constitutional law, and state and federal regulations protecting privacy—much of which predates the internet—are of limited value." (footnote omitted)); Sonia K. Katyal, Note, *Private Accountability in the Age of Artificial Intelligence*, 66 UCLA L. REV. 54, 100 (2019) ("[O]ur existing frameworks for regulating privacy and due process cannot account for the sheer complexity and numerosity of cases of algorithmic discrimination.").

296. *Brnovich v. Democratic Nat'l Comm.*, 594 U.S. 647, 669–71 (2021).

for analyzing the discriminatory effects of a completely different election structure shaped by generative AI. The Court also announced that voters “must tolerate the ‘usual burdens of voting’” and that “[m]ere inconvenience cannot be enough to demonstrate a violation of [Section] 2.”²⁹⁷ Thus, absent evidence of intentional racial discrimination, a secretary of state that moves extended hours of early voting from Thursday to Wednesday because an AI-powered tool determined that doing so would lower turnout among voters likely to cast ballots against the secretary of state—an overwhelming share of whom are people of color—would likely not violate Section 2.²⁹⁸

In light of the judicially imposed requirement of intent to find a violation of the Equal Protection Clause,²⁹⁹ this legal claim would be of limited use in blocking an AI application that disadvantages people of color—such as disproportionately rejecting mail-in ballots cast by voters of color. State courts often “lockstep” their interpretation of state constitutional provisions to mirror federal court interpretations of the U.S. Constitution³⁰⁰ and may also deny relief. “[O]ur existing statutory and constitutional schemes are poorly crafted to address issues of private, algorithmic discrimination”³⁰¹ by AI-powered chatbots, recommendation algorithms, and other tools that are central to modern democratic participation.

Despite the failure of existing law, AI is a social construct shaped by values,³⁰² and the racial harms it produces are not inevitable.³⁰³ Those who design and deploy AI technologies used in the context of democracy directly or indirectly make normative judgments about accuracy, efficiency, equality, and the allocation of political power. Either consciously or through indifference, they make value judgments about the costs of “collateral” harms, such as the proliferation of cultural anxiety and racial polarization, the chilling effects of surveillance on racially marginalized communities, the entrenchment of incumbent politicians and populations, and the acceptable rate of erroneous exclusion of ballots cast by legitimate voters.³⁰⁴

297. *Id.* at 669.

298. *See* Bender, *supra* note 9, at 521.

299. *Village of Arlington Heights v. Metro. Hous. Dev. Corp.*, 429 U.S. 252, 265 (1977) (“Proof of racially discriminatory intent or purpose is required to show a violation of the Equal Protection Clause.”).

300. Joshua A. Douglas, *The Right to Vote Under State Constitutions*, 67 VAND. L. REV. 89, 105–10 (2014).

301. Katyal, *supra* note 295, at 100.

302. *See* Jessica Eaglin, *When Critical Race Theory Enters the Law & Technology Frame*, 26 MICH. J. RACE & L. 151, 155 (2021) (“Technologies are inherently social artifacts.”).

303. Alondra Nelson, *The Right Way to Regulate AI: Focus on Its Possibilities, Not Its Perils*, FOREIGN AFFS. (Jan. 12, 2024), <https://www.foreignaffairs.com/united-states/right-way-regulate-artificial-intelligence-alondra-nelson> (on file with the *Iowa Law Review*) (“Although AI systems are powerful, they remain tools made by humans, and their uses are not preordained. Their effects are not inevitable.”).

304. *Cf.* Jessica M. Eaglin, *Constructing Recidivism Risk*, 67 EMORY L.J. 59, 88–101 (2017) (discussing how certain concerns such as accuracy, equality, and the role of punishment affect normative judgments and societal values).

Those who develop and deploy AI generally make these choices in the absence of law,³⁰⁵ and how (or if) the law responds to shape these normative judgments reflects our values about racially inclusive democracy. Policymakers and citizens should not simply defer to technologists, market forces, and unfettered applications of AI.³⁰⁶ Many technologists lack expertise in navigating racial issues and are not inherently qualified to unilaterally determine the future of democracy in the United States. As Mathias Risse wrote:

Much as technology and democracy are not natural allies, technologists are not natural champions of or even obviously qualified advisers in democracy. . . . Citizens in democracies must not rely on tech experts in hopes that they will make sure technology is used to advance rather than undermine democracy. “Technological advancements must be widely debated in democratic politics, and citizens should take an active interest in these matters.”³⁰⁷

To advance racially inclusive democracy, we will need explicit goals for AI and the law that regulates it.³⁰⁸ In developing such objectives, we should not limit ourselves to existing societal norms like minimizing bias. Instead, we should also aspire to more ambitious goals, such as eliminating racial entrenchment and affirmatively facilitating a well-functioning, racially inclusive, pluralistic democracy.³⁰⁹

Granted, the First Amendment and other doctrines may limit some proposals to regulate AI.³¹⁰ For example, the Supreme Court has held in particular contexts that the First Amendment can protect a deliberate lie,³¹¹ and a law prohibiting all deepfakes would face significant First Amendment challenges.³¹² Many courts have held that computer code is speech protected

305. *Id.* at 99–100.

306. *Cf.* Nelson, *supra* note 303 (“When it comes to technology, policymakers too often believe that their approaches are constrained by a product’s novelty and must be subject to the views of expert creators.”).

307. MATHIAS RISSE, *POLITICAL THEORY OF THE DIGITAL AGE: WHERE ARTIFICIAL INTELLIGENCE MIGHT TAKE US* 71 (Cambridge Univ. Press 2023).

308. *See* Nelson, *supra* note 303.

309. *See* Jessica M. Eaglin, *Racializing Algorithms*, 111 CALIF. L. REV. 753, 757 (2023) (“[C]oncern with racial disparities is not the same as critically questioning race and racial hierarchies in law.”).

310. *See* Eugene Volokh, Mark A. Lemley & Peter Henderson, *Freedom of Speech and AI Output*, 3 J. FREE SPEECH L. 651, 651–58 (2023) (asserting that the generative AI output is protected by the First Amendment because of the rights of the programs’ creators to speak and the users’ rights to listen and speak).

311. *See, e.g.,* *United States v. Alvarez*, 567 U.S. 709, 729–30 (2012) (holding for the first time that the First Amendment protected a deliberate lie uttered with actual malice and distinguishing that from permissible regulation of perjury or impersonating a government officer).

312. Chesney & Citron, *supra* note 9, at 1788–92 (“[I]t is unlikely that a flat ban on deep fakes could withstand constitutional challenge. Deep fakes implicate freedom of expression, even though they involve intentionally false statements.”); *see also* Hasen, *supra* note 9, at 546–53 (acknowledging heightened First Amendment scrutiny of restrictions on deepfakes, and asserting that “a truth-in-labeling regime should be the top way to deal with the deep fake problem”).

by conventional First Amendment doctrines,³¹³ despite the fact that such rulings often undermine First Amendment values and broader democratic principles.³¹⁴

The existence of constitutional concerns, however, should not cause us to retreat from developing regulation of AI to mitigate racial harms to democracy. In other contexts, policymakers tailor laws to advance public policy priorities such as preventing crime and compatible land use, while at the same time respecting constitutional provisions like the Fourth Amendment and the Takings Clause. Further, other entities that can mitigate racial harms to democracy—such as private companies—are not bound by the same constitutional constraints as government.³¹⁵ While constitutional values warrant serious analysis, terms such as “free speech” and “colorblindness” should not be accepted as magic words that shut down debate or attempts to legislate.

A primary function of this Article is to provide a landscape of the broad array of challenges to democracy from AI that extend well beyond deepfakes and to analyze those challenges in the context of race. In light of the various ways in which AI is transforming democracy, no single legal proposal will comprehensively prevent the myriad of existing and potential challenges. Multiple legal solutions and private-sector practices—accompanied by collaboration between government, the private sector, and civil society—are required to address the problems.³¹⁶

As a result, this Part develops broad principles that can be applied in different contexts to help craft laws and private practices tailored to address different types of harms.³¹⁷ All of the principles are applicable to AI regulators,

313. See, e.g., *Universal City Studios, Inc. v. Corley*, 273 F.3d 429, 445–49 (2d Cir. 2001) (analyzing the application of the First Amendment to code, and concluding that “computer programs constructed from code can merit First Amendment protection”); *Junger v. Daley*, 209 F.3d 481, 485 (6th Cir. 2000) (“Because computer source code is an expressive means for the exchange of information and ideas about computer programming, we hold that it is protected by the First Amendment.”).

314. Xiangnong Wang, *De-Coding Free Speech: A First Amendment Theory for the Digital Age*, 2021 WIS. L. REV. 1373, 1389–405 (asserting judicial decisions finding that computer code is speech are ill-suited to modern realities, insulate private actors from accountability, and undermine democratic values, and proposing a judicial framework to evaluate code that advances First Amendment values).

315. See *Moody v. NetChoice, LLC*, 603 U.S. 707, 740–43 (2024) (holding platform content moderation by platforms constitutes expression that receives First Amendment protection, and that a Texas law intended “to correct the mix of speech” on platforms does not pass even a less stringent intermediate form of First Amendment review).

316. GOLDSTEIN ET AL., *supra* note 57, at 4 (“[T]here is no silver bullet that will singularly dismantle the threat of language models in influence operations. . . . Instead, to effectively mitigate the threat, a whole of society approach, marrying multiple mitigations, will likely be necessary.”).

317. While the principles below are tailored to the unique context of race and democracy in the United States, several sources inspired the principles, including but not limited to the following: *Blueprint for an AI Bill of Rights*, WHITE HOUSE, <https://www.whitehouse.gov/ostp/ai-bi-ll-of-rights> [<https://perma.cc/9P74-BCLP>]; LAW’S COMM. FOR C.R. UNDER L., ONLINE CIVIL RIGHTS ACT (2023), <https://www.lawyerscommittee.org/wp-content/uploads/2023/12/LCCRUL-Model-AI-Bill.pdf> [<https://perma.cc/2GF3-ZDL9>]; Statement of Yael Eisenstat, Vice President, Ctr. for Tech. & Soc’y Anti-Defamation League, for U.S. Senate AI Insight Forum: Elections & Democracy, to U.S. Sen. Chuck Schumer et al. (Nov. 3, 2023), <https://www.schumer.senate.gov/imo/media/doc/Yael%20Eisenstat%20-%20Statement.pdf> [<https://perma.cc/CL6M-ER8g>] [hereinafter Statement of Yael Eisenstat]; Katyal, *supra* note 295, at 100; Panditharatne, *supra* note 75.

AI developers, and social media platforms. Two principles, *anticipate racial harms to democracy* and *provide meaningful accountability*, are particularly relevant to those who deploy AI tools. Aspects of the first three principles (*anticipate racial harms to democracy*, *facilitate pluralism*, and *mitigate racial disinformation and manipulation*) can be incorporated as core values into AI model training processes so that they refrain from producing content and engaging in activities that are inconsistent with these principles.³¹⁸

Some may argue that these principles actually facilitate racial harms to democracy by giving actors that purport to commit to the principles an unwarranted veneer of legitimacy, opting instead for the abolition of AI applications in the context of democracy.³¹⁹ I appreciate this perspective, particularly in light of persistent historical and contemporary racial challenges in the United States and the transformative potential of AI.

I do believe, however, that some fair-minded people in government and the private sector desire a racially inclusive, pluralistic democracy and seek direction on how to mitigate racial harms to democracy from AI. While the principles below will not eliminate all racial harms, a serious commitment to them can make things better. Also, the desire for a competitive advantage by both politicians and companies suggests that AI applications will continue to permeate U.S. democracy, and that real change may be more likely to result from equipping potential allies with tools to mitigate the harms of AI rather than calls for the abolition of AI that go unheeded (although calls for abolition may prod companies to take more substantive actions to mitigate racial harms to democracy, including support for government regulation). I also recognize that, over time, technological developments or the consistent failure of politicians and companies to seriously address these issues may warrant a different approach.

A. *ANTICIPATE RACIAL HARMS TO DEMOCRACY*

American democracy should be protected from automated systems that facilitate foreseeable racial harms, including but not limited to racial polarization, psychometric manipulation, racially-targeted deception, vote

318. See *Claude's Constitution*, ANTHROPIC (May 9, 2023), <https://www.anthropic.com/news/claude-constitution> (on file with the *Iowa Law Review*) (describing using human values inspired from sources like the Universal Declaration of Human Rights to train and govern Anthropic's ClaudeAI); Yuntao Bai et al., *Constitutional AI: Harmlessness from AI Feedback 2-5*, 10-16 (Dec. 15, 2022) (unpublished manuscript), <https://arxiv.org/abs/2212.08073> [<https://perma.cc/XC56-ZGBK>] (describing human oversight provided only through human principles (i.e., foundational values), and a multiple-step training process in which AI models train each other to be harmless by comparing alternative outputs and assessing which cause less harm and are most consistent with the original human principles).

319. Cf. Southerland, *supra* note 223, at 2, 6-14, 73-78 (2023) (asserting that commissions designed to enhance community control over police use of technology for surveillance "risk[s] legitimizing surveillance technologies," and concludes with "a vision that looks to an abolitionist horizon"); Okidegbe, *supra* note 172, at 1692-700, 1705-15 (asserting that processes that increase democratic participation in how algorithms are procured and used can build unwarranted trust and legitimacy).

dilution and suppression, and racial entrenchment.³²⁰ AI developers should not design systems with a reasonably foreseeable possibility of facilitating racial harm to democracy, and developers have a duty of care to affirmatively research and understand racial harms to democracy and how their systems could facilitate such harms.³²¹ AI developers should consult with voting rights experts, voter mobilization experts, political scientists, sociologists, campaign strategists, community groups, content moderation experts, and other experts to understand potential racial harms of an automated system, including harms from downstream applications that are outside of the intended use of the system.³²²

Developers should submit systems for pre-deployment testing (including real-world scenarios), and developers should agree to ongoing assessment and independent monitoring and evaluation of a system, its impacts, and actions taken to mitigate racial harms to democracy.³²³ Indeed, policymakers should require that AI developers engage in these steps, and institutional entities that purchase and/or use AI applications or services with direct or indirect applications to democracy should ensure that they work only with vendors and other developers that comply with this principle and others.

For example, for reasons articulated in Section III.D.2., county election officials, federal and state officials, AI developers, and AI vendors should anticipate that particular local election offices—including but not limited to many of those that serve large populations of color—are particularly attractive targets for AI-powered attacks by foreign or domestic actors aiming to shift election outcomes. In light of these potential harms, federal and state officials and the private sector should work to ensure that these local election offices have access to resources to most effectively mitigate targeted threats, such as multifactor authentication, Cybersecurity and Infrastructure Security Agency

320. Cf. *Blueprint for an AI Bill of Rights*, *supra* note 317 (“You should be protected from unsafe or ineffective systems.”).

321. Cf. *Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*, at 12–14, 43–52, 69–71 COM (2021) 206 final (Apr. 21, 2021) (anticipating the risks of particular AI applications, and *banning* applications that pose unacceptable risk (e.g., systems that infer political beliefs, race, and other protected categories; subconscious behavioral manipulation; widespread scraping of facial images for facial recognition databases), *heavily regulating* high-risk applications (e.g., law enforcement and border control), and *requiring transparency* for limited-risk applications (e.g., systems that generate synthetic text, audio, and video)).

322. See GINA M. RAIMONDO & LAURIE E. LOCASCIO, NAT’L INST. OF STANDARDS & TECH., U.S. DEP’T OF COM., ARTIFICIAL INTELLIGENCE RISK MANAGEMENT FRAMEWORK (AI RMF 1.0), at 1–9 (2023), <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf> [<https://perma.cc/L2YD-2PCW>] (recognizing that AI technologies can pose risks that “amplify, perpetuate, or exacerbate inequitable or undesirable outcomes for individuals and communities,” including those that emerge from “the social context in which it is deployed” and providing guidance for organizations developing or using AI systems on identifying and tracking risks—including emerging risks).

323. See Algorithmic Accountability Act of 2023, H.R. 5628, 118th Cong. § 4(a)(4)(E) (2023) (requiring that the Federal Trade Commission mandate impact assessments (including on race) for automated decision systems prior to and after their deployment).

(“CISA”) assistance, training on managing AI threats, and other cyber and physical security protections.³²⁴

Anticipating racial harms to democracy is important because a small group of companies enjoy not only concentrated power in commercial-sphere decision-making shaped by AI, but also significant influence throughout U.S. democracy.³²⁵ These companies have workforces and governance structures that do not represent the diversity of our democracy,³²⁶ and the assets they control—foundation models and the computational infrastructure required so they can continuously perform billions of computations—are expensive and are not democratically distributed.³²⁷

The need to anticipate racial harms to democracy is also evidenced by past failures of the tech industry to do so.³²⁸ Racial bias is well documented in

324. See Norden & Ramachandran, *supra* note 273.

325. See Daniel Stid, “Down the Road Is Coming Much Sooner Than You Think”—Stephanie Bell on the Rise of AI and Civil Society’s Response, ART ASS’N. (May 31, 2023), <https://www.theartofassociation.org/blog/down-the-road-is-coming-much-sooner-than-you-thinkstephanie-bell-on-ai-and-civil-society> [https://perma.cc/G8HA-8L3M] (“On our current path, the advance of AI has a propensity to become extremely inequalitarian, to exclude the folks who are already somewhat marginalized, to concentrate wealth, and with it power, in the hands of a very small number of companies in a very small number of countries run by a very small number of people.”); AGATHE BALAYN & SEDA GÜRSSES, EUR. DIGIT. RTS., BEYOND DEBIASING: REGULATING AI AND ITS INEQUALITIES 108–09 (2021), https://edri.org/wp-content/uploads/2021/09/EDRi_Beyond-Debiasing-Report_Online.pdf [https://perma.cc/8GBH-PM3R] (“[W]e find that research and innovation in machine learning are also becoming concentrated in the hands of these same key players, since their computational resources far outpace what any independent institution can provide. . . . The integration of ML services into computational infrastructures deeply impacts the ability of public and social actors, and especially members of marginalized communities, to question these infrastructures or demand red-lines, prohibitions or other governance responses mitigating the harms of AI systems.”); JULIA RHODES DAVIS, ELIZA MCCULLOUGH, SARAH TREUHAFT & RACHEL GICHINGA, AI NOW, ADVANCING RACIAL EQUITY THROUGH TECHNOLOGY POLICY 1, 8–9 (2023), <https://ainowinstitute.org/wp-content/uploads/2023/09/AINOW-Racial-Equity-Report-Sept-2023.pdf> [https://perma.cc/RH7W-KA7B] (“Ownership of and control over key parts of computational infrastructures give [Amazon, Google, and Microsoft] privileged access to production data, as well as an advantage in shaping machine learning practices. . . . [It] also enables these large tech firms to dominate the provision of data-driven services and to consolidate economic, political, and social power through access to greater amounts of data, more market share, and increased revenue and profits.” (footnote omitted)).

326. BOMMASANI ET AL., *supra* note 51, at 133; Kaitlyn Schwanemann, *Experts Call for More Diversity to Combat Bias in Artificial Intelligence*, CNN (Dec. 15, 2023, 2:07 PM), <https://www.cnn.com/2023/12/15/us/diversity-artificial-intelligence-bias-reaj/index.html> [https://perma.cc/MQ4Z-KTEZ]; Kari Paul, ‘Disastrous’ Lack of Diversity in AI Industry Perpetuates Bias, Study Finds, GUARDIAN (Apr. 16, 2019, 8:47 PM), <https://www.theguardian.com/technology/2019/apr/16/artificial-intelligence-lack-diversity-new-york-university-study> [https://perma.cc/LY43-S6KH].

327. Jonathan Vanian & Kif Leswing, *ChatGPT and Generative AI Are Booming, but the Costs Can Be Extraordinary*, CNBC (Mar. 13, 2023, 2:09 AM), <https://www.cnbc.com/2023/03/13/chatgpt-and-generative-ai-are-booming-but-at-a-very-expensive-price.html> [https://perma.cc/WP6W-3PBD] (estimating that training a large language model like GPT-3 could cost over \$4 million, that even after training costs could be hundreds of thousands of dollars per month to software companies like OpenAI and data-driven services like Amazon Web Services because LLMs do billions of calculations in response to a prompt).

328. BOMMASANI ET AL., *supra* note 51, at 135 (“Machine learning has an established trackrecord [sic] of inequitable impact, with much of the burden of its harms borne by marginalized communities.”).

a wide range of machine learning applications, including predicting an individual's creditworthiness, likelihood of recidivism, and suitability for a job, as well as healthcare diagnosis, facial recognition, and search engines.³²⁹ Photo categorization applications that label Black people as gorillas have gone unfixed for years.³³⁰ Search engines have produced images of children playing games in response to queries for “[w]hite girls,” but pornographic images in response to queries for “Black girls.”³³¹ Many more examples demonstrate that tech industry approaches that fail to be conscious of race-specific harms actually perpetuate racial harms.³³²

B. FACILITATE PLURALISM AND PREVENT ALGORITHMIC DISCRIMINATION

Voters should not face racial discrimination by AI applications in the context of democracy, and AI models should not undermine a healthy, pluralistic democracy by homogenizing viewpoints or excluding voters of color or their legitimate perspectives.³³³

Too often, bias in models and the design of AI models used in chatbots, synthetic content creation tools, translation tools, voice assistance tools, content moderation tools, voter list maintenance processes, and other tools throughout the election process unfairly misrepresent or exclude communities and voters of color.³³⁴ Recognizing that only a handful of foundation models are the basis for a large number of different AI applications, racial biases

329. WEIDINGER ET AL., *supra* note 56, at 11; Julia Angwin, Jeff Larson, Surya Mattu & Lauren Kirchner, *Machine Bias*, PROPUBLICA (May 23, 2016), <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> [<https://perma.cc/82DM-KGEC>]; *see, e.g.*, Dena F. Mujtaba & Nihar R. Mahapatra, *Ethical Considerations in AI-Based Recruitment*, INST. ELEC. & ELECS. ENG'RS 1 (Dec. 23, 2019), <https://ieeexplore.ieee.org/document/8937920> [<https://perma.cc/4VGL-7TT3>] (explaining how AI can have human biases); Ziad Obermeyer, Brian Powers, Christine Vogeli & Sendhil Mullainathan, *Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations*, 366 SCI. 447, 449–53 (2019) (“[Suggesting] that the choice of convenient, seemingly effective proxies for ground truth can be an important source of algorithmic bias in many contexts.”); Joy Buolamwini & Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, 81 PROC. MACH. LEARNING RSCH. 77, 80–90 (2018); SAFIYA UMOJA NOBLE, *ALGORITHMS OF OPPRESSION: HOW SEARCH ENGINES REINFORCE RACISM* 1–14, 64–109 (2018) (“[Demonstrating] that machine learning algorithms can discriminate based on classes like race and gender.”).

330. Nico Grant & Kashmir Hill, *Google's Photo App Still Can't Find Gorillas. And Neither Can Apple's.*, N.Y. TIMES (May 22, 2023), <https://www.nytimes.com/2023/05/22/technology/ai-photo-labels-google-apple.html> (on file with the *Iowa Law Review*); Hewitt, *Congressional Testimony*, *supra* note 228, at 35.

331. NOBLE, *supra* note 329, at 66–68.

332. DAVIS ET AL., *supra* note 325, at 3 (“Current tech regulatory efforts remain race-blind, which allows the tech sector to continue to perpetuate racial equity harms.”).

333. *Cf. Blueprint for an AI Bill of Rights*, *supra* note 317 (“You should not face discrimination by algorithms and systems should be used and designed in an equitable way.”).

334. *See supra* Section III.B.

within a single foundation model can adversely affect hundreds of applications that shape democracy.³³⁵

This is not simply limited to unfair bias; it includes the design of many AI models that scrape the web for data and mimic the content of communities and perspectives that are most visible, mathematically defaulting to averages and dominant patterns. This automation of homogenization is a significant racial harm to a pluralistic, liberal democracy that values respect for the coexistence of diverse interests and viewpoints.³³⁶ It not only prevents self-government and autonomy by many within a racially diverse democracy, but also entrenches the preferences of past democratic majorities. Outputs from the flawed design impair government's ability to most effectively address the current and emerging needs of its increasingly diverse population.

AI developers and deployers should be held to a high standard of care to routinely anticipate these racial harms and mitigate them.³³⁷ This standard of care should include racial equity impact assessments before and after deployment and use of representative data, as well as independent audits and public disclosure of disparities and challenges. AI is evolving rapidly, and the scope of guardrails could change, which makes audits to measure bias, hallucinations, and other mistakes particularly valuable. Human review of technology-enabled election administration flags (e.g., invalid ballot signatures, outdated registrations) to carefully review the merit of such flags and their racial impact is also warranted.

This standard of care should also acknowledge that even representative data can entrench racial stereotypes, racial polarization, and marginalization of communities of color, and it should require proactive design efforts to mitigate the homogenization and exclusion of perspectives of communities of color. The standard of care should also include post-deployment feedback loops to learn from challenges and gaps. Here again, policymakers should require that entities that develop and deploy AI engage in these steps.

C. *MITIGATE RACIAL DISINFORMATION AND MANIPULATION*

American democracy should be protected from AI systems that facilitate racial disinformation and impersonation, psychometric manipulation, polarization, racial hostility and cultural anxiety, deception that deters voting,

335. BOMMASANI ET AL., *supra* note 51, at 131 (“Since the same foundation model serves as the basis for myriad applications, biases in the representation of people propagate to many applications and settings.”).

336. *See generally* Sorensen et al., *supra* note 5; Gordon et al., *supra* note 5; *Value Kaleidoscope*, *supra* note 5. *See also* GLEN WEYL, AUDREY TANG & COMMUNITY, PLURALITY: THE FUTURE OF COLLABORATIVE TECHNOLOGY AND DEMOCRACY 82, 114 (2023), <https://github.com/pluralitybo/ok/plurality/releases/latest/download/Plurality-english.pdf> [<https://perma.cc/V6GA-4PNL>] (defining plurality as “technology for collaboration across social difference,” and noting that “[l]iberal democracies often celebrate themselves as pluralistic societies Yet despite this formal commitment to pluralism and democracy, almost every country has been forced by the limits of available information systems to homogenize and simplify social institutions in a monist atomist mold that runs into direct conflict with such values”).

337. BOMMASANI ET AL., *supra* note 51, at 152.

and surveillance that chills speech. Systems should not collect broad sets of data from various contexts, and then use that data to target particular racial groups or individuals with tailored messages designed to manipulate them, deceive them in political participation, stoke cultural anxiety or racial polarization, or chill them from exercising expressive liberties.

Practices that implement these principles might include legislation that prohibits using generative AI to defraud people about the time, place, and manner of elections,³³⁸ as well as watermarking of synthetic content and content provenance standards.³³⁹ Social media platforms could enact and enforce platform policies on synthetic content (e.g., disclosure), invest in systems that detect synthetic activities, adopt common interoperable digital provenance standards, disclose information on how election-related ads are targeted, and emphasize recognizing and containing racial harms to democracy.³⁴⁰

Another action to implement this principle would be equitable data privacy protections. Generic data privacy proposals often fall short in effectively protecting people of color.³⁴¹ Data privacy protections tailored to be more racially inclusive would help mitigate AI-facilitated racially targeted disinformation, manipulation, surveillance, polarization, and election restrictions that dilute or deny votes from communities of color.

Granted, as with many civil rights proposals, there will be questions about paternalism, “reverse racism,” and borderline cases. For example, should subtle messages designed to reduce documented anti-Black implicit bias in society be prohibited? What about covert messages found to increase voter turnout targeted at Latino, Asian American, and Black communities, which tend to vote at lower rates than whites? Should historical accounts of racism be restricted if they increase hostility or anxiety among some Latino and Asian Americans?

The prospect of grappling with these difficult questions, however, should not prompt us to simply accept that unbridled disinformation, racial polarization, cultural anxiety, psychometric manipulation, and compelled assimilation are inevitable byproducts of a pluralistic society. We should be able to embrace democratic values and implement policies that facilitate information integrity, building new coalitions across racial lines, democratic inclusion, individual autonomy, and pluralism.

338. A provision restricting communications that mislead voters about voting requirements and procedures likely poses few constitutional problems. *See* *Minn. Voters All. v. Mansky*, 585 U.S. 1, 18 n.4 (2018) (“We do not doubt that the State may prohibit messages intended to mislead voters about voting requirements and procedures.”).

339. For a discussion of watermarking and content provenance, see Siddarth Srinivasan, *Detecting AI Fingerprints: A Guide to Watermarking and Beyond*, BROOKINGS (Jan. 4, 2024), <https://www.brookings.edu/articles/detecting-ai-fingerprints-a-guide-to-watermarking-and-beyond> [<https://perma.cc/HG8B-2JXZ>].

340. Thanks to Farbod Faraji of Protect Democracy for exchanges that stimulated some of these thoughts. *See also* Statement of Yael Eisenstat, *supra* note 317.

341. *See, e.g.*, Allen, *supra* note 224, 932–33 (articulating five goals for race-conscious and antiracist privacy and data-protection-law policy reforms).

D. *PROVIDE MEANINGFUL ACCOUNTABILITY*

Those who develop and deploy automated systems should be held accountable when their systems cause reasonably foreseeable racial harms to democracy. Private industry should adopt metrics and standards to ensure accountability, and actually enforce policies they voluntarily enact.³⁴² Government should adopt laws to ensure accountability.

For example, private sector companies (e.g., social media platforms, election administration vendors, AI developers, computational infrastructure providers) currently make significant decisions that determine the racial allocation of political liberties in the United States. The rights to vote, speak, assemble, petition, and generally participate in democracy are at stake when an algorithm makes a decision that mail-in ballots from Asian Americans are more likely to be rejected, that a disproportionately large number of African Americans are to be purged from the voter rolls, and that particular outputs will be produced by a chatbot. As private sector companies focused on innovation, efficiencies, and market share rather than preserving a racially inclusive democracy, their decisions are too often shielded from public view.³⁴³

Lawmakers should consider legal requirements that facilitate more private sector transparency, such as algorithmic racial impact statements,³⁴⁴ racial audits of foundation models and datasets, and enhanced whistleblower protections for current and former AI workers.³⁴⁵ While transparency alone is insufficient, it increases incentives for private companies to be more vigilant.

Lawmakers can also facilitate accountability by providing that future AI regulation can be enforced by the federal government, state attorneys general, and/or private litigants. They could also explicitly state that providers of generative AI do not enjoy immunity from legal liability for content created using their tools.³⁴⁶ Congress could also enhance accountability by appropriating adequate resources to federal agencies to prevent racial harms to democracy. Federal and state lawmakers could regulate the use of AI systems to flag voters to be removed from registration rolls and prevent bots from submitting

342. See Cat Zakrzewski, *ChatGPT Breaks Its Own Rules on Political Messages*, WASH. POST (Aug. 28, 2024, 12:19 PM), <https://www.washingtonpost.com/technology/2023/08/28/ai-2024-election-campaigns-disinformation-ads> [<https://perma.cc/HEC5-SYK5>].

343. Katyal, *supra* note 295, at 99 (“[T]he right to vote . . . and the right to receive information, among others, are all at issue when an algorithm makes its (private) decisions Those decisions are not always subject to public oversight [and] even more problematically, they may be shielded from view, due to trade secrecy and systemic opacity.”). See generally FRANK PASQUALE, *THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION* (Harv. Univ. Press 2016) (exposing how powerful interests abuse secrecy for profit and explaining ways to rein them in).

344. Treasury Bd. of Can. Secretariat, *Directive on Automated Decision-Making*, GOV’T CAN. (Apr. 25, 2023), <https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592> [<https://perma.cc/B5DQ-SXWS>].

345. See generally Katyal, *supra* note 295 (exploring the impending conflict between the protection of civil rights and AI, and focusing on solutions to encourage private corporations to address the issue, such as codes of conduct, impact statements, and whistleblower protections).

346. See LAW’S COMM. FOR C.R. UNDER L., *supra* note 317, at 29–30.

automated challenges to election offices.³⁴⁷ Further, Congress could require that all state and local election officials preclear with federal officials all AI-created maps and voting rules (as well as datasets and algorithms) that apply to elections for federal office before implementation to ensure they are not discriminatory.³⁴⁸

CONCLUSION

AI technologies pose unique opportunities and challenges for the future of racially inclusive democracy in the United States. As our nation becomes more diverse, an explicit goal of AI and the law should be to facilitate our transition to a well-functioning, pluralistic democracy that respects both identity and individual autonomy, facilitates cross-group coalition building, and incentivizes political operatives to respond to demographic change through inclusive representation. Those who develop and deploy AI technologies should mitigate the risks of racial harms to democracy from their applications, and policymakers should ensure that AI regulation contemplates and prevents racial harms to democracy.

347. Panditharatne, *supra* note 75 (“Congress and state legislatures should set baseline requirements governing the official use of AI systems to remove voters from rolls . . . [and] reject the use of bots to transmit automated challenges to voters’ registrations to election offices.”).

348. This *nationwide* preclearance requirement would not run afoul of the U.S. Supreme Court’s opinion that invalidated federal statutory provisions applying preclearance requirements to *select jurisdictions*. *Shelby County v. Holder*, 570 U.S. 529, 556–57 (2013).