# Exploring the Impact of AI on Black Americans:

## Considerations for the Congressional Black Caucus's Policy Initiatives


Nina Dewi Toft Djanegara
Daniel Zhang
Haifa Badi Uz Zaman
Caroline Meinhardt
Gelyn Watkins
Ezinne Nwankwo
Russell Wald
Rohini Kosoglu
Sanmi Koyejo
Michele Elam


BLACK IN AI

Stanford University
Human-Centered
Artificial Intelligence

# Principal Authors

**Nina Dewi Toft Djanegara** is a Ph.D. candidate in anthropology at Stanford University and a Visiting Fellow at the Massachusetts Institute of Technology (MIT). At the time of writing this white paper, she was the associate director of the Technology and Racial Equity Initiative at the Stanford Center for Comparative Studies in Race and Ethnicity (CCSRE). In this role, she oversaw the initiative's practitioner fellows program, graduate and undergraduate fellowships, and strategic development. Her research uses ethnographic and archival methods to explore how computer vision is applied to "solve" political problems. In particular, her dissertation examines how surveillance technology—such as facial recognition and biometric identification—is applied to border management and law enforcement. She holds an MSc in environmental science from Yale University and a BA in international development studies from the University of California, Berkeley.

**Daniel Zhang** is the senior manager for policy initiatives at the Stanford Institute for Human-Centered Artificial Intelligence (HAI), where he leads the institute's policy research, outreach, and education initiatives. With the goal of developing evidence-based AI policy recommendations, his research interests lie at the intersection of technology policy, governance, and societal impact, including translational and original research on AI regulation and standards, the geopolitical implication of emerging technology, and the governance of large-scale ML models. Previously, he was the manager of the AI Index that measured and evaluated the rapid rate of AI advancement. Daniel holds a master's degree in security studies from Georgetown University's Walsh School of Foreign Service, where he concentrated on technology policy, and a bachelor's degree from Furman University.

**Haifa Badi Uz Zaman** is a program manager on the policy and society team at the Stanford Institute for Human-Centered Artificial Intelligence (HAI), where she leads the institute's emerging work on AI in the social sector, and AI in the Global South. Haifa is interested in empowering social sector leaders with the knowledge that they need to more effectively influence AI design, policy, and regulation. At HAI, she also aims to build bridges to connect the institute's work with the needs of Global South countries through research, education, and outreach activities. Haifa has a master's degree in international education policy from Harvard University, and a bachelor's degree in mass communication from the American University of Sharjah. She previously managed research and education programs at the Stanford Cyber Policy Center, the Stanford Center on Philanthropy and Civil Society, Citizen Schools California, and the Aga Khan Development Network.

**Caroline Meinhardt** is the policy research manager at the Stanford Institute for Human-Centered Artificial Intelligence (HAI), where she develops and oversees policy research initiatives. She is passionate about harnessing AI governance research to support the establishment of policies that ensure the safe and responsible development of AI around the world. Prior to joining HAI, Caroline worked as a China-focused consultant and analyst, managing and delivering in-depth research and strategic advice regarding China's development and regulation of emerging technologies including AI. She holds a master's degree in international policy from Stanford University, where her research focused on global governance solutions for AI and other digital and emerging technologies, and a bachelor's degree in Chinese studies from the University of Cambridge.

# Principal Authors (Cont'd)

**Gelyn Watkins** serves as the chief executive officer of Black in AI, a membership and programmatic-based organization focused on broadening representation in the field of AI. She previously served as a consultant with Pivotal Ventures program investment team, where she focused on Responsible AI and Pathways into Venture Capital & Entrepreneurship portfolios within Pivotal's Women & Tech Innovation Strategy. A builder by nature, Gelyn draws from over 20 years of experience across investment management, the startup ecosystem, and the education world as an operator and advisor, leading founders and emerging organizations to their due north. Gelyn's business and strategic acumen intersect with her dedication to guiding founders toward a more equitable future. Shecontinues this strategic- and equity-based work with Black in AI,shepherding Black in AI from its workshop-based roots to an organization with greater capacity and resources to advocate for equity in the field of AI. Gelyn is also the founder of Ala Labs, an incubation lab that gives voice and place to entrepreneurs and their ideas and she holds a Bachelor of Business Administration degree with a minor in Cultural Anthropology from Loyola University Chicago.

**Ezinne Nwankwo** is a Ph.D. candidate in computer science at the University of California at Berkeley. She completed her undergraduate degree at Harvard University and a post-graduate fellowship at the University of Cambridge. Ezinne's research focuses on using statistical and machine learning methods to better understand society (using social data) and to aid in decision-making processes. She is a passionate advocate for students from underrepresented backgrounds.

**Russell Wald** serves as the deputy director for the Stanford Institute for Human-Centered Artificial Intelligence (HAI). In this role he oversees HAI's research, education, communications, administrative activities, industry programs, and policy and society hub. Wald works with HAI's co-directors and faculty leaders to help shape the strategic vision and human-centered mission of HAI. From 2020 - 2022 he served as HAI's first Director of Policy and later Managing Director for Policy and Society. Over the past decade, Wald has held various policy program and government relations positions at Stanford University. He is the co-author of numerous publications on AI, including *Building a National AI Research Resource* (2021) and *Enhancing International Cooperation in AI Research: The Case for a Multilateral AI Research Institute* (2022). Wald is also part of a research team studying potential addictive properties of AI. He is a graduate of UCLA.

**Rohini Kosoglu** is a policy fellow at the Stanford Institute for Human-Centered Artificial Intelligence (HAI). She was recently deputy assistant to the president and domestic policy advisor to the vice president at the White House and chief of staff in the U.S. Senate—the first South Asian woman to serve in these positions. A leading national expert on domestic policy and veteran of the White House, Congress, and presidential campaigns, she has been at the forefront of driving transformative change in social, technology, and economic policy over the last two decades. She is also the director of innovation at the Stanford Byers Center for Biodesign and a venture partner at Fusion Fund, a firm focused on early-stage technology and healthcare investments. She is a former resident fellow at the Harvard Institute of Politics at the Kennedy School and graduated from the University of Michigan and George Washington University.

# Principal Authors (Cont'd)

**Sanmi Koyejo** is an assistant professor of computer science at Stanford University and a faculty affiliate at the Stanford Institute for Human-Centered Artificial Intelligence (HAI). Koyejo leads the Stanford Trustworthy Artificial Intelligence Research (STAIR) group, working to develop the principles and practice of trustworthy machine learning, focusing on applications to neuroscience and healthcare. Koyejo has been the recipient of several awards, including multiple outstanding paper awards, a Skip Ellis early career award, a Sloan fellowship, a Terman faculty fellowship, an NSF CAREER award, a Kavli fellowship, an IJCAI early career spotlight, and a trainee award from the Organization for Human Brain Mapping. Koyejo serves on the Neural Information Processing Systems (NeurIPS) Foundation Board, the Association for Health Learning and Inference (AHLI) Board, and as president of the Black in AI organization.

**Michele Elam** is the William Robertson Coe Professor of Humanities in the English department at Stanford University, a senior fellow of the Institute for Human-Centered Artificial Intelligence (HAI), and a race and technology affiliate at the Center for Comparative Studies in Race and Ethnicity (CCSRE). Former director of African and African American studies, she is also affiliated with the Clayman Institute for Gender Research and with the Wu Tsai Neuroscience Institute. Her research in interdisciplinary humanities connects literature and the social sciences to examine changing cultural interpretations of gender and race. Her work is informed by the understanding that racial perception and identification in particular impacts outcomes for health, wealth, and social justice. More recently, her scholarship examines intersections between race, technology, and the arts. *Race Making in the Age of AI*, her most recent book project, considers how the humanities and arts function as key crucibles through which to frame and address urgent social questions about equity and social justice in socially transformative technologies.

# Acknowledgments

The authors would like to thank David Kyuman Kim, Alaa Youssef, and Benjamin Xie for valuable feedback and Jeanina Casusi, Joe Hinman, Nancy King, Shana Lynch, Carolyn Lehman, and Michi Turner for preparing the publication.

**The Stanford Institute for Human-Centered Artificial Intelligence**
353 Jane Stanford Way, Stanford CA 94305-5008

February 2024, V1.0

# Table of Contents

# Introduction

The Congressional Black Caucus (CBC) has a crucial role to play in the age of artificial intelligence (AI). We understand AI as any computational system that attempts to mimic human intelligence, performing tasks that require learning, reasoning, problem-solving, and decision-making. AI is one of the defining challenges of our time, a technology that holds tremendous promise, but also raises profound questions about our values and our future. While the CBC's policy agenda remains unchanged in the face of the rapid proliferation of AI systems, we believe that it will be crucial for the CBC to apply a new lens to each of its policy focus areas that considers the opportunities and risks of AI development. We hope this paper will serve as a useful resource to help the CBC ground its policy agenda in the context of recent AI developments and their implications for Black Americans.

Sound AI policy must be anchored in a comprehensive and holistic approach that considers the potential for racial biases at every stage of AI development. This includes determining which social problems can be meaningfully addressed by AI, and which decisions are too sensitive to hand over to an algorithm. With this white paper, we also aim to help the CBC develop a thoughtful, forward-looking AI policy strategy that ensures the benefits of this technology are widely shared and its risks are carefully managed.

## The myth of tech neutrality

Technology is never neutral. It reflects and reinforces the values of those who develop it.[1] However, we believe that technology is more than just a container for existing social biases; it is also a tool that can actively contribute to or exacerbate racism. This insight is grounded in the work of scholars like Dorothy Roberts,[2] who documented how scientists have reinforced and redefined common-sense understandings of race throughout history, and Simone Browne,[3] who outlined how surveillance technologies emerged out of the desire to monitor and control Black bodies.[4] Like other technologies that came before it, AI is imbued with social and political values, including biases around race. For example, AI systems have been shown to perpetuate and amplify racial discrimination in employment, housing, and criminal justice.[5] In particular, overreliance on algorithms to make sensitive decisions about loans or hiring can exclude people from financial services or accessing other opportunities—a process known as "algorithmic redlining."[6] A compounding factor is that among the people who research, develop, and invest in such AI systems, relatively few are Black.[7] These examples demonstrate the need for a critical and intentional approach to the design and application of technology, one that prioritizes equity, justice, and human dignity.

While AI holds the potential to deepen racial inequalities, it can also benefit Black communities. If deployed carefully, AI has the power to improve access to healthcare and education, as well as create new economic opportunities. For example, AI can help doctors make more accurate diagnoses and provide personalized treatment plans, particularly in underserved communities where access to healthcare is limited. AI can also assist educators in tailoring lessons to individual student needs, increasing the chances of academic success for all students, including those from low-income and minority communities. Additionally, AI has the potential to redress systemic biases in banking and financial services, promoting greater access to economic opportunities for Black Americans. Our vision for human-centered AI is rooted in the belief that AI

should be assistive, augmenting, and complementing human capabilities but never replacing human judgment.[8] We write this white paper with the conviction that the CBC has more to contribute to AI policy than simply correcting racial biases. Instead, it can help steer AI to ensure the well-being and prosperity of Black communities.

## How do computers see race?

AI tends to see race in restrictive, oversimplified ways that can reinforce racial stereotypes and color lines and/or lead to the mis-categorization of people. AI models conceptualize race in terms of neatly defined and fixed categories, oftentimes relying on the five racial types used by the U.S. Census Bureau.[9] However, racial categories are not clearly delineated or a priori biological types. The Census Bureau's racial classification practices, for example, have historically been informed by political and ideological needs and interests.[10]

The racial categorization imposed by our data collection methods and adopted by AI models also fails to appreciate the cultural and social components of race and how it intersects with other identities, such as gender, class, and sexuality. Many people's social identities resist easy categorization. Consider the difficulty people who are mixed-race or gender-queer will have placing themselves in a single box. As Michele Elam argues, racial categorization based on fixed, static, programmable data points misrepresents—and in some cases misdirects attention from—the important social and political dimensions to racial formation, which go far beyond skin color and physiognomy.[11]

Yet, it is difficult to overcome this limitation of AI because narrow, unidimensional understandings of race are integral to the technology itself. Computer scientists hoping to produce fairer AI systems tend to concentrate their efforts on the model training stage, during which AI can inherit racial biases from historical datasets, operating on the belief that better data can resolve the problem of AI bias[12] As many research has highlighted, racial biases can enter AI at various stages of the technology development life cycle, from problem-setting to deployment.[13] However, the problems at hand extend beyond technical bias or bad data and cannot simply be resolved by diversifying the workforce of computer scientists. To fully grasp the impacts of AI on marginalized communities, it is imperative to recognize how AI models understand and infer race.

## Structure of the white paper

In this paper, we explain recent developments in artificial intelligence that we believe are most relevant for the CBC. First, we discuss the rapid evolution of generative AI models, a breakthrough technology which is finding applications across sectors. Then, we turn to healthcare and education and outline how these sectors are being transformed by AI. Ultimately, this white paper is intended as an educational document, laying out the relevant issues and debates, rather than a set of definitive policy recommendations. It remains the task of policymakers to determine what kinds of regulation will be required to ensure that the significant promises of AI can be realized.

While issues like algorithmically enabled policing and surveillance are important concerns for Black Americans, these topics have been well-documented by other researchers and journalists.[14] Our intent in this white paper is to share information about sectors that complement and potentially expand the CBC's policy platform and are less commonly invoked when talking

about race and AI. In each section, we explain what AI is currently capable of and where it is being used, and then explore the promises and perils of AI in the near future. This guidance will help the CBC take proactive steps to ensure that AI technology is developed and applied in ways that protect civil rights and promote racial justice. Finally, while this paper was inspired by conversations with CBC staff, the insights it puts forward are broadly applicable to other groups that could be marginalized by AI.

# 1. Generative AI models

## Key Takeaways

Generative AI systems have the potential to complement human labor. Yet they also can exacerbate existing barriers and vulnerabilities faced by Black Americans and lower-income, marginalized communities.

Generative AI could, arguably, lower barriers to entry for Black creators and artists, but there is the real risk that it can leave them and their content even more vulnerable to exploitation and appropriation than it has historically.

AI-generated content is eroding information integrity and public trust by becoming increasingly difficult to distinguish from real content. It also reproduces racial and other stereotypes that are harmful to Black and other marginalized communities.

Generative models hold various economic opportunities: They can help boost worker productivity and show particular promise for upskilling workers with lower baseline skill sets. But lack of Black representation across the AI industry contributes to gaps in wealth creation opportunities.

The environmental harms of generative AI tools disproportionately impact already marginalized populations, furthering environmental inequities.

Generative AI models—algorithms that generate original content from simple text prompts—are among the most powerful and transformative technologies of our time. The recent launch of OpenAI's ChatGPT is just the tip of the iceberg of generative AI's potential. Last year, we saw the release of powerful models, including text-to-image (e.g., Stable Diffusion, DALL-E 2), text-to-audio (e.g., Whisper), and text-to-video (e.g., Imagen Video, Make-A-Video), all of which can be adapted to a variety of downstream tasks.

From creating new product designs and enhancing customer experiences through online chatbots to optimizing business processes and assisting medical breakthroughs, generative AI presents a wide range of opportunities for all Americans. Near-future applications for generative models include virtual assistants, design prototyping, and creative content creation. If used responsibly, such AI systems could complement human efforts, making us more productive and creative. However, these models also exhibit factual inaccuracies and harmful stereotypes. As such, they amplify and exacerbate existing societal biases that can lead to increased hostility, discrimination, and violence toward marginalized communities.[15] They also undermine our collective trust in information—harms that are felt most potently by Black Americans and other marginalized communities but can't be addressed using simple technical fixes.

Below, we outline the current capabilities of generative models and analyze the prospective perils and promises associated with this powerful new technology as they relate to creative expression, information integrity, bias and discrimination, economic opportunity, and the environment.

Addressing these issues will require grappling with questions such as: How should artists and creators be compensated for the use of their creative works in generative AI systems? What mechanisms are needed to verify the veracity of and restore public trust in online content? How do we prevent generative AI systems from perpetuating and exacerbating already harmful biases and stereotypes? How can we ensure Black Americans have equal access to the economic opportunities and mobility promised by generative AI? How do we counter AI's emerging impact on environmental inequity?

We address issues specific to applications of generative AI in healthcare and education separately, as these application areas pose a broad range of

challenges and opportunities worth exploring in more detail that go beyond generative AI.

## Creative expression

The outputs of generative models closely resemble the works of intelligent human creativity, and the latest advancements in text-to-image generators have produced visually striking images and videos that have created a stir in both the artistic and AI communities. Within the media and entertainment industries, generative AI has the potential to become a powerful augmentative or assistive technology for creators. For marginalized communities specifically, generative AI could provide a platform for diverse voices to be heard in the creative industries by lowering the barriers to entry for artists from underrepresented groups who would otherwise face discrimination and limited access to resources.[16] In other words, generative models could allow Black creators to convert their ideas and experiences into original content and share them with the world without the need for expensive software or extensive training that may be inaccessible or lacking institutional backing.

However, existing copyright laws and norms around creative production are unable to account for the inputs and outputs of generative models, which gives rise to important questions including: Who owns the intellectual property rights for works created by these models?[17] How can we ensure artists provide consent and receive appropriate acknowledgment and compensation for such works?[18] These questions are still up for debate and are the subject of several ongoing lawsuits.[19] Artists have already raised alarms about AI-generated content that is derivative of their creative labor.[20] Although the generated images are new, these models carry over stylistic features from their training data, which can leave creators from marginalized communities vulnerable to exploitation and appropriation of their content by others who profit from their labor. This must be considered in relation to the history of Black contributions to the arts and the appropriation of Black creative culture—or what Perry A. Hall calls the "virtual 'strip-mining' of Black musical genius and aesthetic innovation."[21] Without clear legal and regulatory frameworks in place to address issues of provenance and ownership, creators are unable to assert their rights or to seek redress for any copyright and intellectual property infringements. Importantly, this can lead to a situation where the works of underrepresented creators are devalued, suppressed, or even erased, perpetuating existing power dynamics that favor dominant cultural narratives.[22]

*Without clear legal and regulatory frameworks in place to address issues of provenance and ownership, creators are unable to assert their rights or to seek redress for any copyright and intellectual property infringements.*

## Information integrity

The erosion of public trust posed by generative AI could have serious consequences for society as a whole. Recent research shows that people are unable to distinguish whether text was written by a human

or an AI, suggesting that our existing heuristics are insufficient for detecting AI authorship.[23] Convincing deepfakes created by powerful generative models can be used for malicious purposes such as spreading disinformation and propaganda or blackmailing individuals.[24] For instance, users have generated convincing images of political figures like Donald Trump, Alexandra Ocasio-Cortez, and Pope Francis that went viral on social media; though these images were framed as humorous parodies, the potential for more dangerous misinformation is clear.[25]

As AI-generated content becomes more prevalent and difficult to distinguish from human-generated content, individuals may become more skeptical and distrustful of the information they receive. Without proper evaluation and authenticity checks, widespread confusion and distrust can lead to a breakdown in communication and collaboration, making it harder for individuals and organizations to work together effectively. Additionally, mistrust in information could lead to unwarranted skepticism about legitimate content distributed by activists (e.g., videos documenting police brutality or other human rights abuses), thereby affecting the capability of civil society to speak truth to power. Furthermore, tools developed to detect deepfakes have been found to be biased, performing best on Caucasian faces and disproportionately outputting incorrect detection results for certain racial groups.[26]

## Bias and discrimination

More broadly, bias and discrimination are a well-documented issue for generative AI systems. These models can pose risks to marginalized communities due to the reproduction of harmful stereotypes.[27] For instance, users of Stable Diffusion have created violent and sexualized images, exacerbating the bias and discrimination of minorities embedded in the models' datasets.[28] The reproduction of racial stereotypes by generative models is not only offensive in and of itself, but it can also result in real-world harms. Studies of implicit bias have shown how associations between images and racial stereotypes contribute to the dehumanization of Black people in criminal justice contexts.[29] In addition, researchers have demonstrated how racial stereotypes can be internalized by people who have already been stigmatized, causing additional psychological distress and undermining their educational and professional outcomes.[30] The erosion of clear delineations between what is real content and what is not, in turn, is making it even more difficult to expose and correct harmful stereotypes.

AI systems are widely known to have exacerbated existing racial biases in financial services, housing, and a variety of public services. For example, research has found that Black and Hispanic borrowers were charged higher rates by an AI-based mortgage lending system than white borrowers applying for the same loans.[31] The proliferation of generative AI tools in a variety of public and private sector decision-making environments and their potential to cause discriminatory outcomes warrants further scrutiny.[32]

## Economic opportunity

Investors are bullish about generative AI's contribution to market productivity, with Goldman Sachs estimating that it could raise the overall global GDP by 7 percent.[33] However, unequal representation of Black Americans and other minorities in the AI industry points to uneven participation in this wealth creation process: In 2018, Black workers represented only 2.5 percent of Google's workforce and 4 percent of Facebook's and Microsoft's.[34] According to a survey conducted by BLCK VC, Black investors make up only

3 percent of the VC industry.[35] While the diversity of computer science (CS) students is increasing in North America, in 2021 only around 4 percent of new CS bachelor's, master's, and PhD graduates were Black or African American.[36] Technological training and upskilling interventions will be crucial in efforts to narrow these economic gaps. Yet researchers have also warned that industry must go beyond improving the AI talent pipeline by addressing more systemic issues that prevent minorities from staying in the field, including exclusionary hiring practices, harassment, unfair compensation, and power asymmetries.[37]

The advent of generative AI is set to restructure the landscape of productivity in various sectors, from automated content creation to data analysis. While these advancements promise a surge in efficiency and the automation of mundane tasks, they also inadvertently risk exacerbating existing racial wealth gaps. Several studies have shown that automation technologies have magnified wage inequality in the United States, driven by relative wage declines for workers specializing in routine tasks.[38] Moreover, Black Americans may face the profound impacts of automation from a notably disadvantaged standpoint due to their higher representation in occupations more susceptible to automation, such as truck drivers, food service personnel, and office clerks.[39] The proliferation of AI could lead to a disproportionate accrual of productivity benefits to majority-owned companies and communities.

However, rather than replacing workers entirely, we believe that with the appropriate intervention, generative models can augment human capabilities and help boost worker productivity. MIT research studying a population of marketers, grant writers, consultants, data analysts, human resource professionals, and managers shows that workers with a lower baseline skill set saw the greatest improvements in productivity.[40] In other words, ChatGPT was able to "upskill" low-ability workers by increasing the quality of their output and allowing them to compete with high performers. Another study demonstrates that the performance of customer support agents using AI tools to guide their conversations improved by 14 percent on average and more than 30 percent for the least experienced workers.[41] This could have transformative potential for those who have historically had fewer opportunities to gain experience and training, making them more productive and competitive members of the workforce.

*... rather than replacing workers entirely, we believe that with the appropriate intervention, generative models can augment human capabilities and help boost worker productivity.*

## Environmental impact

The negative environmental impact of generative AI development has increasingly come into focus as these AI systems have become more widely accessible. Training such AI systems requires enormous computing power and, consequently, a vast number of energy-hungry servers. Researchers estimate that the process of training an AI model can emit more than 626,000 pounds of carbon dioxide equivalent, an

amount nearly five times what an average American car emits during its lifetime.[42] For example, the final training of the powerful open-access large language model BLOOM is estimated to have emitted up to 24.7 tonnes of carbon dioxide equivalent.[43] Yet training accounts for only a fraction of a model's carbon footprint. In the case of BLOOM, if you account for carbon emissions from other processes, ranging from the manufacturing of hardware (including semiconductors) to the energy consumption of other operational processes, the model's carbon footprint more than doubles.[44] Water consumption is another important environmental factor: Training GPT-3 in Microsoft's U.S.-based data centers is said to have directly evaporated 700,000 liters of clean freshwater.[45]

This is particularly concerning for marginalized communities, which—as environmental racism literature has widely documented—are often the first to be impacted by resulting harms. Scholars have already documented AI's emerging environmental inequity, highlighting that AI's environmental footprint is disproportionately higher in certain regions.[46] For example, data centers located in areas with higher outside temperatures (e.g., drought-stricken Arizona) have less efficient on-site water consumption processes (required for cooling and for off-site electricity generation), leading to the surrounding areas being more negatively impacted by the environmental toll of AI development.

# 2. Medicine and Healthcare

## Key Takeaways

AI holds particular promise for healthcare applications, but questions remain regarding the safety and equity of medical algorithms and AI-assisted medical devices.

Medical imaging and diagnostics is an area where AI already excels at reducing unnecessary deaths, but employing diverse training datasets will be crucial to ensuring equal performance across racial groups.

AI's ability to enable more effective, personalized medical treatment plans promises to reduce racial disparities in healthcare, but limited availability of such bespoke medicine could also widen those same disparities.

While the use of AI-powered calculation tools can help allocate scarce healthcare resources to those most in need, they have also been found to show racial biases and prioritize cost reduction at the expense of patient needs.

Healthcare represents one of the most promising application areas for artificial intelligence. For instance, AI models can assist doctors and medical practitioners in diagnosing patients, screening medical records to identify which people have the greatest risk of developing cancer, or providing customized treatment plans unique to individual patients. AI is also being used to research understudied and rare diseases, such as sarcoidosis and sickle cell anemia, that occur at higher frequencies among Black Americans.[47] Human-centered AI can speed up certain tasks so doctors are able to devote more of their time to the important aspects of healthcare that cannot be automated, like communicating with patients and understanding their concerns. Overall, we believe that medical AI should scaffold the work of doctors, not serve as a replacement for care.

For decades, doctors have used medical algorithms to guide decision-making. However, these rule-based algorithms took the form of flowcharts, decision trees, scoring systems, and scientific formulas, which could be calculated by hand or through basic computation. In contrast, contemporary medical AI models are highly complex and leverage large amounts of data to discover unseen patterns or make sophisticated predictions. Whereas conventional medical algorithms followed step-by-step rules and were based on knowledge from clinical practice, medical AI models are often a "black box," even for their developers who may be unable to pinpoint how or why a model outputs certain decisions.

While studies have shown that AI holds particular promise for healthcare applications, translating academic research to products that are safe for deployment will require input from regulators. In particular, there are significant concerns about the safety of medical AI devices. A 2021 survey of FDA-approved medical AI devices revealed that 97 percent of the devices had only been evaluated using retrospective data; they were not tested on live patients.[48] Many of the devices were only tested on limited populations clustered around a few geographic sites, and therefore their performance may not be broadly generalizable.

Commercial medical AI, sometimes referred to as Software as a Medical Device (SaMD), is currently less regulated than pharmaceuticals, and the nature of these algorithms will present different challenges to regulators and policymakers.[49] In comparison to pharmaceuticals or traditional medical

*While studies have shown that AI holds particular promise for healthcare applications, translating academic research to products that are safe for deployment will require input from regulators.*

devices, SaMDs are unique in that their performance will change significantly over time due to the fact that "learning algorithms" continue to evolve as they are introduced to more data—a concern that was acknowledged by the FDA commissioner in 2019.[50] Furthermore, whereas new drugs need to be tested in extensive clinical trials, there is no such requirement for AI-powered medical devices.

More broadly, policymakers will also need to grapple with the reality that many patients remain reluctant to make use of healthcare products and services that are powered by medical AI.[51] The COVID-19 pandemic exacerbated a pre-existing distrust of the medical system at large, especially among Black and other minority communities, as the pandemic highlighted a variety of racial inequities evident in aspects ranging from infection and mortality rates to the vaccine rollout.[52] Amid increasing reports of racial bias in medical AI, Black communities are likely to be among those most resistant to embracing AI-assisted healthcare solutions.[53]

To move ahead, Congress must consider the following questions: Which medical decisions should be assisted by AI, and which will require more significant human oversight? How should medical algorithms be evaluated to ensure equity across demographics?[54] How often should SaMDs be re-evaluated, and what long-term monitoring efforts will be necessary? How can policymakers and healthcare providers foster trust in medical AI solutions? Which agencies and institutions should be tasked with independent oversight of medical AI? As a guide to help address these questions, we provide an overview of some of the central debates in medical AI and highlight their relevance to people of color.

## Diagnostics and medical imaging

AI is particularly skilled in *pattern recognition*, or spotting phenomena that repeat across images, including patterns that are not easily discernible to the human eye. This capability is particularly useful when applied to medical imaging. When given X-rays, mammograms, or brain scans, AI can spot indicators like blood clots or tumors, helping doctors diagnose patients more effectively and efficiently. For instance, an NYU study demonstrated that AI enhanced the performance of human radiologists.[55] When doctors leveraged AI tools to analyze breast ultrasounds, they were more effective at detecting breast cancer compared to either humans or computers working alone. They also reduced the number of false positives, thereby decreasing the number of unnecessary biopsies. If applied judiciously, AI can also contribute to reducing unnecessary deaths due to delayed diagnosis and medical error. However, to ensure fairness and racial equity, it will also be important to ensure that AI models are trained on sufficiently diverse datasets and that performance is generalizable across racial groups.[56]

## Precision medicine

Imagine a world in which every patient is given a unique treatment plan that takes into account their specific genetic traits, medical history, environmental exposure, and social circumstances. This is the future imagined by advocates of "precision medicine." These personalized recommendations will be enabled by AI, which is capable of digesting large amounts of data and comparing it against known cases. For instance, AI models could draw upon data from wearable devices and individual biomarkers to provide up-to-date recommendations that are custom-made for each person. In principle, this would make treatments more effective and reduce adverse reactions while also lowering healthcare costs. Proponents also argue that precision medicine will reduce racial disparities in healthcare because treatment will be tailored to the individual rather than relying on race-adjusted algorithms.[57] Though the use of racial categories is common in clinical decision-making, it can cause medical professionals to make false assumptions about a person's genetic background and/or predisposition to specific diseases, thereby leading to serious medical errors.[58]

However, some remain concerned that bespoke medicine will only be available to a select few, therefore widening the gap between those who receive high-quality medical care and those who do not.[59] Finally, there is a risk that disproportionate emphasis on medical AI can take attention away from larger structural factors that affect patient outcomes. To realize the promise of precision medicine, health policy must also address the social and environmental determinants of health that can affect a patient's ability to follow an AI-customized treatment plan.

*These cases illustrate the need for regulation to ensure that Black Americans are not unduly affected by healthcare decisions that prioritize cost reduction at the expense of patient need.*

## Operational management

Alongside patient care, AI is being used to manage other aspects of the healthcare industry. Here, we make a distinction between medicine as a scientific practice and healthcare as a business that considers factors like cost, efficiency, and resource availability. AI can be used to make operational decisions like scheduling patients and personnel or assigning hospital beds to those who are most in need. This can be especially helpful when allocating scarce resources, like distributing vaccines to those who are most at risk or those who could contribute most to flattening the curve.[60]

However, Stanford researchers have demonstrated that the use of AI-powered calculation tools has a significant impact on the pricing of healthcare.[61] More troubling, another team of researchers evaluated a widely used AI risk assessment tool and found evidence of racial bias.[62] This predictive tool was used by health insurance companies to identify which patients may need additional care to defray more expensive treatment costs down the road. However,

evaluators found that the algorithm underestimated the needs of Black patients; it assigned lower risk scores to Black patients, even when they were sicker than white patients who received the same score. While the algorithm explicitly eliminated race as an input, racial bias resulted from the decision to use the cost of care as a proxy for the severity of need. Since the algorithm drew upon past patient data, it reproduced historical patterns of racial bias, namely the tendency for the healthcare system to spend more money on treating white patients than their Black counterparts.[63] Yet, if the AI tool were recalibrated to take into account frequency and severity of chronic illness, it would have flagged more than twice as many Black patients as candidates for early intervention. These cases illustrate the need for regulation to ensure that Black Americans are not unduly affected by healthcare decisions that prioritize cost reduction at the expense of patient need.[64]

# 3. Education

## Key Takeaways

AI will fundamentally disrupt education as we know it, requiring us to reconsider assessment norms. But it also may improve learning outcomes for students at under-resourced schools and increase access to high-quality education.

AI-enabled adaptive learning tools and assistive devices may help students whose learning needs are not met in conventional classroom settings, increase student engagement, and enable teachers to focus more on personalized student assessments.

Predictive AI tools that forecast student performance could allow early interventions that help low-achieving students, but biased predictions could backfire. College recruitment and admissions algorithms may also discriminate against certain demographics.

AI-powered video analytics and behavioral biometrics could perpetuate inequalities by performing worse for darker-skinned people and acting as surveillance tools.

Since the 2022 release of ChatGPT and other generative text models, a major concern has been how these models will disrupt norms around assessment. AI, like other innovations that have come before it (e.g., the calculator or the computer), requires societies to rethink what skills are valuable and, in turn, what knowledge is worth learning. Therefore, our education system will need to reckon with the role of generative models in the classroom.

Many teachers may need to reconsider the purpose and format of written assignments and reconfigure essay questions to assess the kinds of critical thinking and evaluation skills that cannot be easily mimicked by generative AI. It will become even more crucial than before to focus on developing valuable 21st century competencies such as critical thinking, evaluation, and problem-solving. In particular, new assessment norms or pedagogical approaches that prioritize these skills must utilize a participatory approach through meaningful co-design.[65] Early feedback from educators and underserved communities will ensure that AI technologies are designed with the needs of diverse student populations in mind. For example, generative AI tools that utilize computer-generated dialogue can maximize student learning outcomes if they incorporate the sociocultural and linguistic context of diverse users.[66] This is essential in the context of the CBC's existing policy agenda that emphasizes equitable access to quality education.

In order to promote such learning, policy-makers should consider questions such as: How can we leverage AI advancements to better serve students from underrepresented and under-resourced backgrounds? What baseline technical resources will schools need to work with AI? What structural conditions need to be met to ensure that AI is accessible, experienced, and built in a manner that promotes equity and serves the interests of every student?

Our focus in this section is not AI literacy or AI education, though we acknowledge that gaining these competencies will be important to prepare students for the changing professional and civic landscapes.[67] We recognize that addressing educational equity requires an ecosystem-wide approach that extends beyond formal education to after-school programs that focus on upskilling underserved students. While a detailed analysis of

this topic is outside the scope of this white paper, it is worth mentioning that greater public investment in afterschool STEM programs serving Black communities can strengthen the Black talent pipeline into AI-relevant careers.[68]

In this white paper, we concentrate instead on the potential for AI to improve equity and access to high-quality education while at the same time outlining the possibilities for harm. Prior history and research tells us that new technologies often exacerbate educational inequities.[69] We must, therefore, engage with historical and structural inequities within and beyond the educational system as AI tools including generative AI are increasingly used to advance educational equity.[70]

*...policymakers must remain wary of the implications of introducing external devices into classrooms that risk making underserved students even more vulnerable to the whims and exploitative data collection practices of large technology corporations.*

## Bridging achievement gaps

Arguably, AI-enabled adaptive learning tools can better serve students and bridge achievement gaps by providing tailored lesson plans and assignments.[71] AI-powered learning devices would be adaptable; for instance, a student struggling with a particular concept would see it repeated in future problem sets, while a student who has already mastered that skill would be accelerated to receive new material. This personalization would be especially impactful for students whose needs are not met in conventional classroom settings—for instance, those with learning disabilities, non-native speakers, and neurodivergent or highly gifted students. AI-enabled assistive devices can also help students with special learning needs, such as those with autism, by providing behavioral interventions for better learning outcomes.[72] When determining the best pathway for rolling out AI-enabled devices in schools, policymakers may look to lessons learned from previous implementations of education technology, such as laptops or tablets.[73] In particular, policymakers must remain wary of the implications of introducing external devices into classrooms that risk making underserved students even more vulnerable to the whims and exploitative data collection practices of large technology corporations.[74]

In addition, educators are hopeful that AI learning devices will increase student engagement and enthusiasm for learning since they would create dynamic modes of interaction and adapt to student performance and interests. AI-enabled lesson plans may use principles of game design, ensuring that they capture students' attention and reward success in a way that motivates students to continue learning.[75] Other researchers have shown that AI-powered feedback tools can significantly increase teachers' uptake of student ideas and improve students' learning experience, as well as their optimism about their

academic future.[76] However, the efficacy of tools may not generalize across all groups of students.[77] Finally, AI could assist teachers with the time-consuming work of grading at scale and provide more personalized assessments of students' individual strengths and weaknesses.[78] This form of assessment would be based on measurable indicators, thereby ensuring that students are gaining the necessary skills while still being more flexible and student-centered in comparison to the one-size-fits-all approach of standardized testing.[79]

In particular, generative AI that can help students and teachers generate text, images, and other media is seen as a promising tool to support under-resourced students and schools. As a tool that offers currently affordable, near-instant, seemingly infinite generation of media, generative AI creates new opportunities for adaptable learning of creativity, critical thinking, and other 21st century skills across disciplines.[80] However, for-profit organizations use datasets that reflect historical biases to develop these generative AI tools. This creates risks of homogenization and assimilation of language and culture, where students may be required to use unfamiliar language to use the tools effectively, and they may encounter AI-generated output that contains harmful stereotypes.[81] Incorporating generative AI in education settings also raises concerns regarding protecting students' privacy in K-12 and higher education.[82]

## Prediction and risk assessment

Another proposed use of AI for education relates to forecasting and prediction of student performance. Risk-assessment algorithms could, in principle, identify students who are struggling academically or exhibit signs of psychological distress. Such "early warning systems" may be able to intervene and help low-achieving students before they drop out of school.[83] However, any predictions of student success will necessarily draw upon past data, and will likely encode historical biases that have made certain students more likely to thrive than others. This should cause policymakers to seriously question the ethics of using such AI tools in classrooms.

Like automated sentencing algorithms that attempt to forecast a person's likelihood of recidivism, the predictions made by these algorithms are not guaranteed to come true—and research has shown that inaccurate predictions can expose people to harm.[84] For instance, students who are predicted to do poorly in school may be stigmatized by teachers, excluded from scholarship opportunities, or funneled into lower-track classes. Students themselves may internalize these predictions, become disheartened, and exert less effort in school. In fact, research has shown the importance of fostering a "growth mindset" and inculcating the belief that one's abilities are not set in stone; AI predictions of student success may run counter to these efforts.[85] Regulation may be needed to place guardrails around the use of student data (e.g., demographic data) for predictive AI in schools, bearing in mind students' ability to surpass expectations.[86] Moreover, policymakers and school leaders must address existing structural issues in schools around teacher readiness and training before rushing to adopt new technologies that may not be set up for success in implementation.

Within higher education, intelligent matching algorithms are also being used in recruitment and admissions to determine which potential applicants

*When considering deployment, it will require setting clear guidelines about data collection and storage in a way that respects students' privacy and self-autonomy.*

should be targeted, which applications should be accepted, and who should be awarded scholarships. Intelligent matching algorithms help universities meet enrollment metrics but are not necessarily conducive to student success. A Brookings report found that enrollment algorithms use predictions about likelihood to enroll and willingness to pay as factors in deciding how to allocate scholarships; applicants who are deemed more likely to attend a given university may actually be offered less money.[87] Depending on how these algorithms are calibrated, they could inadvertently discriminate against certain demographics, or be used to target a more diverse student population.

## Classroom monitoring

Video analytics are another form of artificial intelligence used in educational spaces. Video analytics refers to systems that use AI to detect objects, movements, and patterns within video footage. In classroom settings, video analytics and behavioral biometrics might be used to monitor which students are present or absent, or whether students appear to the instructor to be paying attention.[88]

During the COVID-19 pandemic, as many schools transitioned to remote instruction, similar systems were used for identity verification (e.g., exam proctoring software that employs facial recognition to ensure that the correct person is taking an exam). However, facial recognition has been shown to be less effective for darker-skinned people, and some students reported being locked out of important exams.[89] Others expressed that they felt surveilled by automated eye-tracking software that purports to detect whether students are cheating on exams.[90]

Some scholars have noted how AI learning tools can be considered as methods of surveillance; policymakers must therefore proceed with deep caution to ensure that the use of these technologies in schools does not exacerbate the "school to prison pipeline."[91] This should entail a frank discussion of when—if ever—deploying such monitoring tools is ethical. When considering deployment, it will require setting clear guidelines about data collection and storage in a way that respects students' privacy and self-autonomy. Should students have rights to opt out of classroom monitoring and, if so, how can those rights be exercised? Who has ownership of AI models built from student data? How will these systems be audited to ensure their intrusiveness is proportional to net gains in learning outcomes?

# Conclusion

This white paper highlights the duality of promises and challenges that rapid AI development and adoption poses for Black communities in the United States. Moving beyond important but already well-documented concerns related to algorithmically enabled policing and surveillance, we show how AI holds both the risk of deepening and the potential to reduce racial inequities in three crucial areas:

1. **Generative AI** systems could lower barriers to entry in the media and entertainment industries and help bolster the capabilities of lower-skilled workers. However, such systems also make Black creators and artists even more vulnerable to exploitation and have already been shown to reproduce harmful racial stereotypes and perpetuate environmental inequities.

2. In **healthcare**, AI-powered devices and resource allocation software, if carefully deployed, could enable the lowering of medical costs, personalized medical treatment plans, and a more equitable allocation of resources for Black Americans and others who are often overlooked by the healthcare system. Yet the very same tools could widen disparities by encoding racial biases, prioritizing cost reduction at the expense of patient needs, and limiting access to bespoke healthcare services.

3. AI tools employed in **educational** settings to assist teachers and students could help bridge achievement gaps by improving the learning outcomes of students in under-resourced schools. At the same time, such systems could exacerbate discrimination against certain minority demographics—especially in-classroom video analytics and behavioral biometrics tools that act as a form of surveillance and perform worse for darker-skinned people.

Given the vast potential impact of AI in these and many more areas, the CBC should develop an AI policy strategy that tackles the complex implications of AI for ongoing efforts to eliminate racial inequalities. The CBC has a unique opportunity to help steer the development and regulation of AI at a critical time to ensure that Black Americans' needs and concerns are reflected in relevant government initiatives, to enable Black communities to benefit from AI progress, and to prevent the widening of racial inequities through biased AI tools.

The three areas we highlight should be viewed as a starting point. Of course, there are many other areas—ranging from the already-mentioned criminal justice system to financial services, housing, climate change, and public administration—in which the adoption of AI systems presents concerns and opportunities. In its efforts to formulate an approach to AI, the CBC must consider the civil rights and racial justice implications of these areas holistically in order to help steer AI development in a direction that ensures the well-being and prosperity of Black communities.

# Endnotes

1 Langdon Winner, "Do Artifacts Have Politics?" *Daedalus* 109, no. 1 (Winter 1980): 121-36, http://www.jstor.org/stable/20024652.
2 Dorothy Roberts, *Fatal Invention: How Science, Politics, and Big Business Re-Create Race in the Twenty-First Century* (New York: The New Press, 2012).
3 Simone Browne, *Dark Matters: On the Surveillance of Blackness* (Durham: Duke University Press, 2015).
4 Our analysis and recommendations are grounded in rigorous scholarship, much of which has been carried out by Black academics. In response to the call of the Cite Black Women movement, we take the time to note pioneering Black researchers by name. See Christen A. Smith et al., "Cite Black Women: A Critical Praxis (a statement)," *Feminist Anthropology* 2, No. 1 (March 2021): 10-17, https://anthrosource.onlinelibrary.wiley.com/doi/epdf/10.1002/fea2.12040.
5 See Kathy O'Neil, *Weapons of Math Destruction: How big data increases inequality and threatens democracy. Broadway Books* (New York: Crown Books, 2016); Khari Johnson, "Feds Warn Employers Against Discriminatory Hiring Algorithms," *Wired*, May 16, 2022, https://www.wired.com/story/ai-hiring-bias-doj-eecc-guidance; Nadiyah J. Humber, "A Home for Digital Equity: Algorithmic Redlining and Property Technology," *California Law Review* 111 (October 2023), https://www.californialawreview.org/print/a-home-for-digital-equity; Sonja B. Starr, "Evidence-Based Sentencing and the Scientific Rationalization of Discrimination," *Stanford Law Review* 66, No. 4 (April 2014), 803-72,https://www.jstor.org/stable/24246717.
6 Humber, "A Home for Digital Equity."
7 Karen Hao, "AI's White Guy Problem Isn't Going Away," *MIT Technology Review*, April 17, 2019, https://www.technologyreview.com/2019/04/17/136072/ais-white-guy-problem-isnt-going-away/
8 Hope Reese, "A Human-Centered Approach to the AI Revolution," Stanford Institute for Human-Centered AI, October 17, 2022, https://hai.stanford.edu/news/human-centered-approach-ai-revolution.
9 Morgan Klaus Scheuerman et al., "How We've Taught Algorithms to See Identity: Constructing Race and Gender in Image Databases for Facial Analysis," *Proceedings of the ACM on Human-Computer Interaction* 4, No. 58 (May 29, 2020): 1-35, https://doi.org/10.1145/3392866; United States Census Bureau, About the Topic of Race, https://www.census.gov/topics/population/race/about.html.
10 Jennifer L. Hochschild, "Racial Reorganization and the United States Census 1850-1930: Mulattoes, Half-Breeds, Mixed Parentage, Hindoos, and the Mexican Race," Studies in American Political Development 22, No. 1 (Spring 2008): 59-96, https://scholar.harvard.edu/jlhochschild/publications/racial-reorganization-and-united-states-census-1850-1930-mulattoes-half-br
11 Michele Elam, "Signs Taken for Wonders: AI, Art & the Matter of Race," Daedalus 151, No. 2 (Spring 2022): 198-217, https://doi.org/10.1162/daed_a_01910.
12 Joy Buolamwini and Timnit Gebru, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification," *Proceedings of Machine Learning Research* 81 (2018): 77-91; Keith Kirkpatrick, "It's Not the Algorithm, It's the Data," *Communications of the ACM* 60, No. 2 (January 23, 2017): 21-23, https://doi.org/10.1145/3022181
13 Ruha Benjamin, *Race After Technology: Abolitionist Tools for the New Jim Code* (Polity, 2019); Irene Y. Chen et al., "Ethical Machine Learning in Healthcare," *Annual Review of Biomedical Data Science* (July 2021): 124-44, https://www.annualreviews.org/doi/full/10.1146/annurev-biodatasci-092820-114757; Leonardo Nicoletti and Dina Bass, "Humans are Biased. Generative AI is Even Worse," Bloomberg, 2023, https://www.bloomberg.com/graphics/2023-generative-ai-bias/.
14 See Julia Angwin et al., "Machine Bias," *ProPublica*, May 23, 2016, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing; Andrew D. Selbst, "Disparate Impact in Big Data Policing," *Georgia Law Review* 52, No. 1 (February 2018), https://georgialawreview.org/article/3373-disparate-impact-in-big-data-policing; Jameson Spivack, "Cop Out: Automation in the Criminal Legal System," Georgetown Law Center for Privacy & Technology, March 29, 2023, https://copout.tech/; Kashmir Hill, "Wrongfully Accused by an Algorithm," *New York Times*, June 24, 2020, https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html; Richard A. Berk, "Artificial Intelligence, Predictive Policing, and Risk Assessment for Law Enforcement," *Annual Review of Criminology* 4 (January 2021): 209-37, https://www.annualreviews.org/doi/abs/10.1146/annurev-criminol-051520-012342.
15 Federico Bianchi et al., "Easily Accessible Text-to-Image Generation Amplifies Demographic Stereotypes at Large Scale," *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (June 2023): 1493-1504, https://dl.acm.org/doi/10.1145/3593013.3594095.; Emily Bender et al., "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (March 2021): 610–623, https://doi.org/10.1145/3442188.3445922.
16 Matteo Wong, "The Dawn of Artificial Imagination," *Atlantic*, December 14, 2022, https://www.theatlantic.com/technology/archive/2022/12/generative-ai-technology-human-creativity-imagination/672460/.
17 Christopher T. Zirpoli and Legislative Attorney, "Generative Artificial Intelligence and Copyright Law," Congressional Research Service, September 29, 2023, https://crsreports.congress.gov/product/pdf/LSB/LSB10922.
18 "Core Principles for Artificial Intelligence Applications," Human Artistry Campaign, https://www.humanartistrycampaign.com/.
19 See Blake Brittain, "Getty Images Lawsuit Says Stability AI Misused Photos to Train AI," *Reuters*, February 6, 2023, https://www.reuters.com/legal/getty-images-lawsuit-says-stability-ai-misused-photos-train-ai-2023-02-06/; Timothy B. Lee, "Stable Diffusion Copyright Lawsuits Could Be a Legal Earthquake for AI," *Ars Technica*, April 3, 2023, https://arstechnica.com/tech-policy/2023/04/stable-diffusion-copyright-lawsuits-could-be-a-legal-earthquake-for-ai/.
20 See Maham Javaid, "The Magic Avatar You Paid $3.99 For Is Probably Stolen, Artists Say," *Washington Post*, December 9, 2022, https://www.washingtonpost.com/technology/2022/12/09/lensa-apps-magic-avatars-ai-stolen-data-compromised-ethics/; Melissa Heikkilä, "This Artist Is Dominating AI-Generated Art. And He's Not Happy About It," *MIT Technology Review*, September 16, 2022, https://www.technologyreview.com/2022/09/16/1059598/this-artist-is-dominating-ai-generated-art-and-hes-not-happy-about-it/.
21 See Toni Lester, "Blurred Lines: Where Copyright Ends and Cultural Appropriation Begins: The Case of Robin Thicke versus Bridgeport Music and the Estate of Marvin Gaye," *Hastings Communications and Entertainment Law Journal* 36, No. 2 (2014): 217, https://repository.uclawsf.edu/hastings_comm_ent_law_journal/vol36/iss2/1/; Bruce Ziff and Pratima V. Rao, eds., *Borrowed Power: Essays on Cultural Appropriation* (New Brunswick: Rutgers University Press, 1997).
22 Megan McCluskey, "These TikTok Creators Say They're Still Being Suppressed for Posting Black Lives Matter Content," *Time*, July 22, 2020, https://time.com/5863350/tiktok-black-creators/.
23 See Maurice Jakesch, Jeffrey T. Hancock, and Mor Naaman, "Human Heuristics for AI-Generated Language Are Flawed," *Proceedings of the National Academy of Sciences* 120, No.11 (March 2023), https://www.pnas.org/doi/10.1073/pnas.2208839120; Prabha Kanna, "Was This Written by a Human or AI?," Stanford Institute for Human-Centered AI, March 16, 2023, https://hai.stanford.edu/news/was-written-human-or-ai-tsu.
24 Ben Buchanan et al., "Truth, Lies, and Automation: How Language Models Could Change Disinformation," Center for Security and Emerging Technology, May 2021, https://cset.georgetown.edu/publication/truth-lies-and-automation/.
25 Isaac Stanley-Becker and Drew Harwell, "How a Tiny Company with Few Rules Is Making Fake Images Go Mainstream," *Washington Post*, March 30, 2023, https://www.washingtonpost.com/technology/2023/03/30/midjourney-ai-image-generation-rules/.
26 See Loc Trinh and Yan Liu, "An Examination of Fairness of AI Models for Deepfake Detection," Preprint, submitted May 2, 2021, https://arxiv.org/abs/2105.00558; Ying Xu et al., , "A Comprehensive Analysis of AI Biases in DeepFake Detection with Massively Annotated Databases," Preprint, submitted August 11, 2022, https://arxiv.org/abs/2208.05845.
27 Bianchi et al., "Easily Accessible Text-to-Image Generation Amplifies Demographic Stereotypes at Large Scale."
28 See Dina Bass, "Text-to-Image Tools Make Cool Art but Can Conjure NSFW Pictures, Too," *Bloomberg*, October 14, 2022, https://www.bloomberg.com/news/articles/2022-10-14/artificial-intelligence-makes-cool-art-but-can-conjure-sexist-pictures-too; Ryan Steed and Aylin Caliskan, "Image Representations Learned with Unsupervised Pre-Training Contain Human-Like Biases," *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (March 2021): 701-713, https://dl.acm.org/doi/abs/10.1145/3442188.3445932; Abeba Birhane, Vinay Uday Prabhu, and Emmanuel Kahembwe, "Multimodal Datasets: Misogyny, Pornography, and Malignant Stereotypes," *arXiv*, October 5, 2021, https://arxiv.org/abs/2110.01963.
29 Phillip Atiba Goff et al., "Not Yet Human: Implicit Knowledge, Historical Dehumanization, and Contemporary Consequences," *Journal of Personality and Social Psychology* 94, No. 2 (2008): 292-306, https://psycnet-apa-org.stanford.idm.oclc.org/fulltext/2008-00466-008.html.
30 Steven J. Spencer, Christine Logel, and Paul G. Davies, "Stereotype Threat," *Annual Review of Psychology* 67 (2016): 415-437, https://www.annualreviews.org/doi/abs/10.1146/annurev-psych-073115-103235.
31 Charlton McIlwain, "AI Has Exacerbated Racial Bias in Housing. Could It Help Eliminate It Instead?" *MIT Technology Review*, October 20, 2020, https://www.technologyreview.com/2020/10/20/1009452/ai-has-exacerbated-racial-bias-in-housing-could-it-help-eliminate-it-instead/.
32 See Cade Metz and Adam Satariano, "An Algorithm That Grants Freedom, or Takes It Away," *New York Times*, February 6, 2020, https://www.nytimes.com/2020/02/06/technology/predictive-algorithms-crime.html; Ellora Thadaney Israni, "When an Algorithm Helps Send You to Prison," *New York Times*, October 26, 2017, https://www.nytimes.com/2017/10/26/opinion/algorithm-compas-sentencing-bias.html.
33 Goldman Sachs, "Generative AI Could Raise Global GDP by 7%," April 5, 2023, https://www.goldmansachs.com/intelligence/pages/generative-ai-could-raise-global-gdp-by-7-percent.html.
34 Sarah Myers West, "Discriminating Systems: Gender, Race, and Power in AI–Report," AI Now Institute, April 1, 2019, https://ainowinstitute.org/publication/discriminating-systems-gender-race-and-power-in-ai-2.
35 Porter Braswell, "Black Founders Only Receive 1.4% VC Funds: Here's How to Change That," *Fast Company*, August 9, 2022, https://www.fastcompany.com/90777034/black-founders-only-receive-1-4-vc-funds-heres-how-to-change-that.
36 Nestor Maslej et al., "The 2023 AI Index Report," Stanford Institute for Human-Centered AI, April 2023 https://aiindex.stanford.edu/report/.
37 West, "Discriminating Systems."
38 See Daron Acemoglu and Pascual Restrepo, "Tasks, Automation, and the Rise in US Wage Inequality," *Econometrica* 90, No. 5 (June 2021): 1973-2016, https://www.nber.org/papers/w28920; John Van Reenen, "Wage Inequality, Technology and Trade: 21st Century Evidence," *Labour Economics* 18, No. 6 (December 2011): 730-41, https://www.sciencedirect.com/science/article/abs/pii/S0927537111000613.

39 David Baboolall et al., "Automation and the Future of the African American Workforce," McKinsey & Company, November 14, 2018, https://www.mckinsey.com/featured-insights/future-of-work/automation-and-the-future-of-the-african-american-workforce.
40 Shakked Noy and Whitney Zhang, "Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence," *Science* 381, No. 6654 (July 13, 2023): 187-92, https://www.science.org/doi/10.1126/science.adh2586.
41 Erik Brynjolfsson, Danielle Li, and Lindsey R. Raymond, "Generative AI at Work" (working paper, National Bureau of Economic Research, April 2023), https://www.nber.org/papers/w31161.
42 Emma Strubell, Ananya Ganesh, and Andrew McCallum, "Energy and Policy Considerations for Deep Learning in NLP," Preprint, submitted June 5, 2019, https://arxiv.org/abs/1906.02243.
43 Alexandra Sasha Luccioni, Sylvain Viguier, and Anne-Laure Ligozat, "Estimating the Carbon Footprint of Bloom, a 176b Parameter Language Model," *Journal of Machine Learning Research* 24, No. 253 (2023): 1-15, https://www.jmlr.org/papers/v24/23-0069.html.
44 Luccioni, Viguier, and Ligozat, "Estimating the Carbon Footprint of Bloom."
45 Pengfei Li et al., "Making AI Less 'Thirsty': Uncovering and Addressing the Secret Water Footprint of AI Models," Preprint, submitted April 6, 2023, https://arxiv.org/abs/2304.03271.
46 Pengfei Li et al., "Towards Environmentally Equitable AI via Geographical Load Balancing," UC Riverside, June 20, 2023, https://escholarship.org/uc/item/79c880vf.
47 Ben Solomon et al., "Intelligent Detection and Diagnosis of Rare Diseases: A Case for AI," *Science*, July 15, 2021, https://www.science.org/content/webinar/intelligent-detection-and-diagnosis-rare-diseases-case-ai.
48 Eric Wu et al., "How Medical AI Devices Are Evaluated: Limitations and Recommendations from an Analysis of FDA Approvals," *Nature Medicine* 27, No. 4 (2021): 582-84, https://www.nature.com/articles/s41591-021-01312-x.
49 See U.S. Food and Drug Administration, "Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan," January 2021, https://www.fda.gov/media/145022/download; Katharine Miller, "When AI Reads Medical Images: Regulating to Get It Right," Stanford Institute for Human-Centered AI, December 3, 2020, https://hai.stanford.edu/news/when-ai-reads-medical-images-regulating-get-it-right; David B. Larson, Daniel L. Rubin, and Curtis P. Langlotz, "Improving AI Software for Healthcare Diagnostics," Stanford Institute for Human-Centered AI, July 2021, https://hai.stanford.edu/policy-brief-improving-ai-software-healthcare-diagnostics.
50 U.S. Food and Drug Administration, "Statement from FDA Commissioner Scott Gottlieb, M.D. on Steps Toward a New, Tailored Review Framework for Artificial Intelligence-Based Medical Devices," April 2, 2019, https://www.fda.gov/news-events/press-announcements/statement-fda-commissioner-scott-gottlieb-md-steps-toward-new-tailored-review-framework-artificial.
51 Chiara Longoni and Carey K. Morewedge, "AI Can Outperform Doctors. So Why Don't Patients Trust It?" *Harvard Business Review*, October 30, 2019, https://hbr.org/2019/10/ai-can-outperform-doctors-so-why-dont-patients-trust-it.
52 See Richard A. Oppel Jr. et al., "The Fullest Look Yet at the Racial Inequity of Coronavirus," *New York Times*, July 5, 2020, https://www.nytimes.com/interactive/2020/07/05/us/coronavirus-latinos-african-americans-cdc-data.html; Amy Schoenfeld Walker et al., "Pandemic's Racial Disparities Persist in Vaccine Rollout," *New York Times*, March 5, 2021, https://www.nytimes.com/interactive/2021/03/05/us/vaccine-racial-disparities.html.
53 See Heidi Ledford, "Millions of Black People Affected by Racial Bias in Health-Care Algorithms," *Nature*, October 24, 2019, https://www.nature.com/articles/d41586-019-03228-6; Health Equity, "Racial Bias in Health Care Artificial Intelligence," NIHCM Foundation, September 30, 2021, https://nihcm.org/publications/artificial-intelligences-racial-bias-in-health-care.
54 Xiaoxuan Liu et al., "The Medical Algorithmic Audit," *The Lancet Digital Health* 4, No. 5 (May 2022): e384-97, https://www.sciencedirect.com/science/article/pii/S2589750022000036.
55 Yiqiu Shen et al., "Artificial Intelligence System Reduces False-Positive Findings in the Interpretation of breast Ultrasound Exams," *Nature Communications* 12, No. 1 (2021): 5645, https://www.nature.com/articles/s41467-021-26023-2.
56 See Amit Kaushal, Russ Altman, and Curtis Langlotz, "Toward Fairness in Health Care Training Data," Stanford Institute for Human-Centered AI, October 2020, https://hai.stanford.edu/policy-brief-toward-fairness-health-care-training-data; Edmund L. Andrews, "Are Medical AI Devices Evaluated Appropriately?" Stanford Institute for Human-Centered AI, April 19, 2021, https://hai.stanford.edu/news/are-medical-ai-devices-evaluated-appropriately.
57 Ekaterina Pesheva, "Precision-Driven Health Equity: Can Precision Medicine Reduce Bias and Disparities in Health Care?," Harvard Medical School, September 16, 2021, https://hms.harvard.edu/news/precision-driven-health-equity.
58 See Darshali A. Vyas, Leo G. Eisenstein, and David S. Jones, "Hidden in Plain Sight: Reconsidering the Use of Race Correction in Clinical Algorithms," *New England Journal of Medicine* 383, No. 9 (2020): 874-82, https://www.nejm.org/doi/full/10.1056/NEJMms2004740; Lundy Braun et al., "Racial Categories in Medical Practice: How Useful Are They?," *PLoS Medicine* 4, No. 9 (2007): e271, https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.0040271#s4.
59 Katharine Miller, "AI + Health: How to Prioritize Humans," Stanford Institute for Human-Centered AI, November 16, 2021, https://hai.stanford.edu/news/ai-health-how-prioritize-humans.
60 Jonathan Greig, "How AI Is Being Used for COVID-19 Vaccine Creation and Distribution," TechRepublic, April 20, 2021, https://www.techrepublic.com/article/how-ai-is-being-used-for-covid-19-vaccine-creation-and-distribution/.
61 Anna Zink, Thomas G. McGuire, and Sherri Rose, "Balancing Fairness and Efficiency in Health Plan Payments," Stanford Institute for Human-Centered AI, November 2022, https://hai.stanford.edu/policy-brief-balancing-fairness-and-efficiency-health-plan-payments.
62 Ziad Obermeyer et al., "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations," *Science* 366, No. 6464 (2019): 447-53, https://www.science.org/doi/10.1126/science.aax2342.
63 Carolyn Y. Johnson, "Racial Bias in a Medical Algorithm Favors White Patients over Sicker Black Patients," *Washington Post*, October 24, 2019, https://www.washingtonpost.com/health/2019/10/24/racial-bias-medical-algorithm-favors-white-patients-over-sicker-black-patients/.
64 Casey Ross and Bob Herman, "Denied by AI: How Medicare Advantage Plans Use Algorithms to Cut Off Care for Seniors in Need," *Stat News*, March 13, 2023, https://www.statnews.com/2023/03/13/medicare-advantage-plans-denial-artificial-intelligence/.
65 Kenneth Holstein and Shayan Doroudi, "Equity and Artificial Intelligence in Education: Will 'AIEd' Amplify or Alleviate Inequities in Education?," Prepint, submitted April 27, 2021, https://arxiv.org/abs/2104.12920.
66 Stanford Graduate School of Education, "Stanford Faculty Weigh In on ChatGPT's Shake-Up in Education," December 20, 2022, https://ed.stanford.edu/news/stanford-faculty-weigh-new-ai-chatbot-s-shake-learning-and-teaching.
67 Stanford University, CRAFT AI Literacy Resources, https://craft.stanford.edu/.
68 Matthew Aldeman and Jeritt Williams, "An After-School STEM Program with a Novel Equitable and Inclusive Structure," *Proceedings of the American Society for Engineering Education 2023 Annual Conference*, June 1, 2023, https://par.nsf.gov/biblio/10424309.
69 Justin Reich, *Failure to Disrupt: Why Technology Alone Can't Transform Education* (Cambridge: Harvard University Press, 2020).
70 See Michael Madaio et al., "Beyond 'Fairness': Structural (In)justice Lenses on AI for Education," Preprint, submitted May 18, 2021, https://arxiv.org/abs/2105.08847; Fengchun Miao and Wayne Holmes, "Guidance for Generative AI in Education and Research," UNESCO, 2023, https://unesdoc.unesco.org/ark:/48223/pf0000386693.
71 Rose Luckin et al., *Intelligence Unleashed: An Argument for AI in Education* (London: Pearson Education, 2016).
72 Erin Digitale, "Google Glass Helps Kids with Autism Read Facial Expressions," Stanford Medicine, August 2, 2018, https://med.stanford.edu/news/all-news/2018/08/google-glass-helps-kids-with-autism-read-facial-expressions.html.
73 See Denise K. Frazier, J. Tolbert, and Amber L. Hall, "K4-Teacher Perceptions of a 1:1 Chromebook Rollout during Pandemic Teaching," Preprint, submitted 2021, https://www.researchgate.net/profile/Joshua-Tolbert-2/publication/354961410_K-4_Teacher_Perceptions_of_a_11_Chromebook_Rollout_During_Pandemic_Teaching/links/6155d4884a82eb7cb5d7e8aa/K-4-Teacher-Perceptions-of-a-11-Chromebook-Rollout-During-Pandemic-Teaching.pdf; Alexandra J. Lamb and Jennie Miles Weiner, "Institutional Factors in iPad Rollout, Adoption, and Implementation: Isomorphism and the Case of the Los Angeles Unified School District's iPad Initiative," *International Journal of Education in Mathematics, Science and Technology* 6, No. 2 (2018): 136-54, https://eric.ed.gov/?id=EJ1178351.
74 Neda Atanasoski and Kalindi Vora, *Surrogate Humanity: Race, Robots, and the Politics of Technological Futures* (Durham: Duke University Press, 2019), chap. 4.
75 Dirk Ifenthaler and Scott Joseph Warren, *Advances in Game-Based Learning* (New York: Springer, 2015-2021), https://www.springer.com/series/13094.
76 Dorottya Demszky and Jing Liu, "M-Powering Teachers: Natural Language Processing Powered Feedback Improves 1:1 Instruction and Student Outcomes," *Proceedings of the Tenth ACM Conference on Learning @ Scale*, July 2023, 59-69, https://dl.acm.org/doi/abs/10.1145/3573051.3593379.
77 Shamya Karumbaiah, Jaclyn Ocumpaugh, and Ryan S. Baker, "Context Matters: Differing Implications of Motivation and Help-Seeking in Educational Technology," *International Journal of Artificial Intelligence in Education* (2021): 1-40, https://shamya.github.io/files/IJAIED_FATE_RMDemog.pdf.
78 Mike Wu et al., "ProtoTransformer: A Meta-Learning Approach to Providing Student Feedback," Preprint, submitted July 23, 2021, https://arxiv.org/abs/2107.14035.
79 Daniel Tanner, "Race to the Top and Leave the Children Behind," *Journal of Curriculum Studies* 45, No. 1 (2013): 4-15, https://www.tandfonline.com/doi/full/10.1080/00220272.2012.754946.
80 Wikipedia, "21st Century Skills," https://en.wikipedia.org/wiki/21st_century_skills.
81 Su Lin Blodgett and Michael Madaio, "Risks of AI Foundation Models in Education," Preprint, submittedOctober 19, 2021, https://arxiv.org/abs/2110.10024.
82 See Anna Merod, "Ed Tech Experts Urge Caution on ChatGPT's Student Data Privacy," *K-12 Dive*, March 29, 2023, https://www.k12dive.com/news/chatgpt-student-data-privacy-concern/646297/; WCET, "Welcome to the Wild, Wild West of AI and the Higher Education Institution," May 11, 2023, https://wcet.wiche.edu/frontiers/2023/05/11/welcome-to-the-wild-wild-west-of-ai-and-the-higher-education-institution/.
83 Raheela Asif et al., "Analyzing Undergraduate Students' Performance Using Educational Data Mining,: *Computers & Education* 113 (2017): 177-94, https://www.sciencedirect.com/science/article/abs/pii/S0360131517301124.

84 See Jeff Larson et al., "How We Analyzed the COMPAS Recidivism Algorithm," *ProPublica*, May 23, 2016, https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm; Juan C. Perdomo et al., "Difficult Lessons on Social Prediction from Wisconsin Public Schools," Preprint, submitted April 13, 2023, https://arxiv.org/abs/2304.06205; Todd Feathers, "Takeaways from Our Investigation into Wisconsin's Racially Inequitable Dropout Algorithm," *The Markup*, April 27, 2023, https://themarkup.org/the-breakdown/2023/04/27/takeaways-from-our-investigation-into-wisconsins-racially-inequitable-dropout-algorithm.
85 Aaron Hochanadel and Dora Finamore, "Fixed and Growth Mindset in Education and How Grit Helps Students Persist in the Face of Adversity," *Journal of International Education Research* 11, No. 1 (2015): 47-50, https://www.clutejournals.com/index.php/JIER/article/view/9099.
86 Ryan S. Baker et al., "Using Demographic Data as Predictor Variables: A Questionable Choice," *Journal of Educational Data Mining* 15, No. 2 (2023): 22-52, https://jedm.educationaldatamining.org/index.php/JEDM/article/view/619.
87 Alex Engler, "Enrollment Algorithms Are Contributing to the Crises of Higher Education," Brookings, September 14, 2021, https://www.brookings.edu/research/enrollment-algorithms-are-contributing-to-the-crises-of-higher-education/.
88 Javier Hernandez-Ortega et al., "edBB: Biometrics and Behavior for Assessing Remote Education," Preprint, submitted December 10, 2019, https://arxiv.org/abs/1912.04786.
89 Buolamwini and Gebru, "Gender Shades": Intersectional Accuracy Disparities in Commercial Gender Classification," *Proceedings of Machine Learning Research* 81 (2018): 77-91, https://proceedings.mlr.press/v81/buolamwini18a.html; Jack Morse, "Online Testing Is a Biased Mess, and Senators Are Demanding Answers," *Mashable*, December 3, 2020, https://mashable.com/article/senate-open-letter-remote-proctoring-examsoft-bias-student-privacy.
90 Todd Feathers, "Students Are Rebelling Against Eye-Tracking Exam Surveillance Tools," *Vice*, September 24, 2020, https://www.vice.com/en/article/n7wxvd/students-are-rebelling-against-eye-tracking-exam-surveillance-tools.
91 See Roy D. Pea et al, "Four Surveillance Technologies Creating Challenges for Education," *Learning: Designing the Future* (New York: Springer, 2023), 317, https://link.springer.com/chapter/10.1007/978-3-031-09687-7_19; Damien M. Sojoyner, *First Strike: Educational Enclosures in Black Los Angeles* (Minneapolis: University of Minnesota Press, 2016).

BLACK IN AI

Stanford University
Human-Centered
Artificial Intelligence