

Shreesa Shrestha

Professor Filipa Calado

Foundations of Information, Info 601-07

13th December 2024

Book Review: *Unmasking AI* by Dr. Joy Buolamwini

In the rapidly evolving landscape of artificial intelligence and technological ethics, Dr. Joy Buolamwini's memoir *Unmasking AI: My mission to protect what is human in a world of machines* emerges as a groundbreaking narrative that advances the discourse on technological bias and social justice. As a scientist, activist, and poet of code, Buolamwini uses her voice to advocate for algorithmic justice and ethical technological development, with a particular focus on uplifting the marginalized communities often overlooked in technological innovation. Her narrative is deeply personal, drawing from her experiences as a black woman navigating computer science, making her critique both accessible and emotionally resonant while broadening the appeal of her message. Through her book *Unmasking AI*, Buolamwini urges reimagining technology as a potential tool for social justice and equitable progress. (Buolamwini)

The power of her book lies in its ability to humanize technological critique. By sharing her personal experiences with algorithmic bias, such as facial recognition systems failing to detect her face unless she wore a white mask, Buolamwini bridges the gap between abstract technological critiques and real-world implications (xiii). She also coined the term "Coded gaze" to describe the embedded biases in algorithmic systems that reflect the values and prejudice of their creator and the data used to train them —a concept that forms the central theme of the book.(Buolamwini xiii). Her storytelling

approach sets her work apart from traditional technological analysis, making the complexities of AI bias both understandable and urgent.

Her work stands in dialogue with contemporary works like Dr. Ruha Benjamin's *Race After Technology* and Dr. Shoshana Zuboff's "Big Other: Surveillance Capitalism and the Prospects of an Information Civilization". While Benjamin provides a powerful theoretical analysis of technological racism, Buolamwini offers a transformative approach that combines personal narrative, rigorous scientific research, and advocacy efforts. Buolamwini provides empirical evidence through her work "Gender Shades" discussed in her book, setting her work apart with its actionable insights. Unlike Zuboff's broader examination of the technological power structure and economic mechanisms of digital surveillance, Buolamwini demonstrates how marginalized communities can shape technological narratives.

Early Years: Childhood and Initial Technological Encounters

Buolamwini's journey began during her formative years when her multicultural background and curiosity about technology began to take root. Growing up as a child of Ghanaian immigrants in the United States, she was uniquely positioned to observe technological systems from an outsider's perspective. Her interactions with technological systems that struggled to detect her darker skin were not only mere inconveniences but also catalysts for questioning the biases in these technologies. One of the earliest examples of her encounter can be found in her childhood photographs taken outside studio lights that only showed the whites of her eyes and teeth and not the rest of her facial features (Buolamwini 26). These encounters revealed the systemic blind spots in technological design and made her determined to challenge them. Her experience as an outsider to a predominantly white and male tech industry shaped her

perspective with a commitment to ensuring that technology serves all people equitably (Buolamwini 212).

Education and Technological Awakening

Her academic journey at Georgia Tech and the MIT Media Lab was not a straightforward trajectory with technology but a path of continuous challenge and purposeful redirection. At Georgia Tech, Buolamwini leveraged technology to address real-world problems such as her mobile surveying tool to combat neglected tropical diseases in Ethiopia. Later, she joined MIT Media Lab's Center for Civic Media to embrace a role as a creator of futuristic technology (Buolamwini 12). Although she started with a distinct research agenda, her work took an unexpected turn towards her now-famous Gender Studies Project, as a result of personal frustration when facial recognition software failed to recognize her face unless she wore a white mask and experienced the coded gaze (Buolamwini 19, 20). Unlike many technological narratives that start with immediate success, her story is characterized by persistent questioning and a refusal to accept the systemic limitations. Buolamwini's encounter reflects the bias embedded in facial recognition technology that researchers and technologists had either overlooked or ignored.

Exposing Algorithmic Bias

Joy Buolamwini's *Gender Shades* research provides a compelling example of how AI systems can exhibit intersectional bias. The research was part of her Master's thesis at MIT. In the study, she focuses on analyzing facial recognition technologies developed by prominent tech companies like Microsoft, IBM, and Face++. Her methodology for the research is particularly noteworthy. Through her Pilot Parliaments Benchmark, she deliberately selected 1270 parliament members from three African countries—Rwanda,

Senegal, South Africa—and three European countries—Iceland, Finland, and Sweden. This diverse selection was intended to showcase how gender and skin tone intersect in these systems, as well as to highlight how AI can perpetuate or even exacerbate existing societal biases (Buolamwini 98,99, 104).

Acknowledging the limitations of existing research methodologies, Buolamwini developed a nuanced classification system based on skin tone and gender. While constrained by the available datasets to a binary gender classification, she critically emphasized the subjective nature of such classifications. This decision highlights the challenges AI research faces in incorporating the full spectrum of gender identities. Using the gender and skin tone classification, she creates four intersectional groups: Dark-skinned females, dark-skinned males, light-skinned females, and light-skinned males. Buolamwini notes that “classifying people by necessity would be shaped by subjective human choices” (94) emphasizing the fundamental principle of her book: technological systems are not objective, but are shaped by human choices, social conditions, and power relationships.

The study’s findings were striking and unsettling. The AI systems demonstrated significant disparities across skin tone and gender, with the highest error rate for dark-skinned females (up to 34.7% misclassification) and the lowest for light-skinned males (Less than 1% error) (Buolamwini 131). The results are not just statistics but a wake-up call for the tech industry to develop more inclusive and representative datasets for fairer and more accurate systems.

Buolamwini’s research extends far beyond facial recognition, revealing that algorithmic biases are, in fact, a reflection of the social landscapes from which they emerge. Her work exposes how machine learning models when trained on historically

underrepresented datasets, perpetuate systemic discrimination across multiple domains. The intersection of skin tone and gender creates a heightened vulnerability to algorithmic bias. In healthcare sectors, AI-driven diagnostic tools and predictive algorithms have repeatedly demonstrated significant racial and gender disparities, often underdiagnosing or misdiagnosing conditions for black, Indigenous, and people of color (Buolamwini 259, 260). These systematic biases can have life-threatening consequences. AI-powered hiring algorithms have been shown to discriminate against candidates based on their gender (Buolamwini 64). Surveillance technologies used by law-enforced agencies powered by AI have been found to disproportionately target black, brown, and immigrant communities (Buolamwini 62-63).

By carefully documenting these disparities, Buolamwini is doing more than critiquing existing technological systems, she is challenging the tech industry to view inclusivity and diversity not as optional add-ons, but as fundamental prerequisites for ethical technological innovation. Her Gender Shades research is a powerful example of how scientific research can be a tool for social transformation offering concrete pathways to more equitable technological futures. We can see this in the response by IBM when informed about the disparity in their data, they took a proactive stance to improve their facial recognition model. Their improved model performed at 96.5% accuracy on darker-skinned females which is a significant improvement from their previous 65.3% accuracy (Buolamwini 143). Through this example, Buolamwini is showing that meaningful change is possible with the right actions.

Advocacy and Systemic Change

Recognizing that research alone is not sufficient, Buolamwini founded the Algorithmic Justice League (AJL) while at MIT. It is fascinating to see the AJL evolve

from a simple idea into a fully operational organization throughout the book. Through AJL, Buolamwini is dedicated to highlighting and mitigating the social harms of Artificial Intelligence and advocating for more equitable and accountable AI systems. AJL empowers people to challenge biased AI systems and demand ethical practices in technology development. Her research at MIT uncovered significant algorithmic biases, inspired by the belief that “against tyranny, oppression, and erasure, we can choose to resist and offer pathways to liberation” (Buolamwini 37,48, 73). AJL represents a resistance league against such technological tyranny challenging them to create systems that recognize and respect human diversity.

Buolamwini’s journey from academic researcher to global thought leader reflects her core mission of reimagining technology as a potential tool for social justice and equitable progress. Her work quickly transcended academic circles. The trajectory of her advocacy began with the TEDxBeaconStreet talk in 2016, where she shared her MFA exhibit on the Coded Gaze (Buolamwini 33). This marked the beginning of her broader mission to challenge systemic bias. By 2019, her reach expanded to become the face of Olay’s #DecodeTheBias campaign, a beauty campaign developed to increase the number of women in STEM while “creating just, responsible, and inclusive consumer AI products; and elevate public awareness about issues of algorithmic bias”(Buolamwini 247). Whether through beauty ad campaigns or international speaking engagements, she consistently sought to broaden the reach of her message on algorithm bias beyond academic and tech circles. By leveraging the visibility of #DecodeTheBias and TEDTalk, she highlights the potential for collaboration between activists, technologists, and corporations to challenge and reshape biased systems so that they prioritize fairness, diversity, and social impact.

The height of her advocacy was marked by her congressional testimonies and her participation in the 2023 White House discussion on AI. These were not just opportunities to share her insights but deliberate efforts to redefine how technology is developed and implemented. By presenting empirical evidence, Buolamwini urged policymakers and technologists to prioritize fairness, transparency, and inclusivity. (Buolamwini 205, 284)

The power of Buolamwini lies in her multifaceted approach to rigorous academic research, activism, corporate engagement, and policy intervention. Her work shows that reimaging technology goes beyond identifying problems; it involves actively designing and implementing solutions that promote equitable technology and innovation. By 2023, Buolamwini has become more than just a researcher or activist, she is a global thought leader in AI technology ethics. Her TED Talk, powerful congressional testimonies, and international speaking engagements transformed the conversation of AI technology from a narrow technological discourse to a broader discussion about social justice, human rights, and technological accountability. (Buolamwini)

Critical Analysis

The book's primary strength lies in its intersectional approach to examining technological impacts on marginalized communities. She provides concrete examples across multiple domains—facial recognition, hiring algorithms, and healthcare diagnostics—demonstrating how seemingly neutral technologies can perpetuate systemic discrimination. Unlike most critical works, Buolamwini does not just highlight problems but offers constructive solutions to move forward. She proposed strategies such as developing more equitable AI, diverse representation in tech, better testing protocols, and ethical guidelines for AI development. Her narrative successfully

connects theoretical technological analysis to real-world consequences revealing how artificial intelligence can reinforce societal disparities. Buolamwini's work offers a unique approach to technological critique by its blend of personal narrative and academic rigor. Her methodology of presenting personal experiences with scholarly research transforms complex technological concepts into accessible and emotionally resonant discourse.

Unmasking AI stands as a thought-provoking and valuable resource for anyone interested in the social, economic, and political aspects of artificial intelligence, however, it is not without limitations. The most notable challenge lies in the treatment of technical depth. While compelling, the narrative occasionally sacrifices technical nuances for accessibility. The complex algorithmic systems she critiques need more comprehensive examination. The ethical exploration in the books, though passionate and important, feels constrained by its scope. The book raises critical questions about bias, privacy, and technological accountability, but these discussions lack a detailed unpacking of the issues. While Buolamwini provides powerful critique, the ethical dilemmas she presents can feel somewhat one-sided, with a sense of profound problem identification and limited practical resolutions.

Moreover, the book's perspective remains Western-centric with a limitation in its global relevance. Despite addressing technological challenges that have worldwide implications, the discussion of AI ethics, policy, and regulation remains centered in the Western technological and cultural context. Although she acknowledges briefly that AI harms are happening in developing countries as well and her work for now has focused on the Western nations (Buolamwini 278), the engagement of other parts of the world

would have enriched the narrative, offering insights into how different cultures and governments are conceptualizing and engaging with AI.

Perhaps the most significant missed opportunity is the unbalanced approach to AI's potential. While Buolamwini's critique of technological bias is both necessary and powerful, the narrative could have been strengthened by a more nuanced exploration of AI's transformative potential, particularly in educating general readers. AI is not only about facial recognition, employee hiring, or surveillance; its direct and indirect impacts today are immense. AI has revolutionized navigation with real-time traffic updates, instant rerouting, and predictive traffic analysis to reduce bottlenecks (Sunday). It has enhanced weather forecasting to address challenges posed by climate change and extreme weather events (AI Weather). AI has also improved supply chain management by optimizing delivery routes to reduce unnecessary carbon emissions (Burnham), among many other applications. If a person with minimal knowledge of AI reads this book, they might come away thinking that there is nothing positive happening with AI and may choose to distance themselves from it, unaware of how it is already positively impacting their daily lives.

These critiques do not diminish the fundamental importance of the book *Unmasking AI* by Buolamwini which challenges readers to think critically about the technological systems increasingly shaping our world. Her ability to blend personal narrative with rigorous research provides a unique lens to understand the complex relationship between technology, social justice, and humans.

Conclusion

Buolamwini's book is a powerful reminder that, as a member of a generation inheriting complex technological systems, we are not passive recipients but active

creators. It inspires readers to understand that every line of code and every algorithmic decision carry moral weight. The book serves as a crucial call to action for technologists, policymakers, and anyone concerned with the ethical development of artificial intelligence. By revealing the hidden biases embedded in our technological systems through her research, Buolamwini invites readers to imagine and work towards a more equitable future. Her work reminds us that technology is not autonomous but a reflection of our values, our limitations as well as our potential for social transformation. Through her diverse efforts, she has stayed true to her fundamental beliefs: technology can be a strong tool when designed with inclusivity, empathy, and a commitment to humans.

Works Cited

- "AI Weather Forecasting Guide." *Climavision*,
<https://climavision.com/resources/ai-weather-forecasting-guide/#:~:text=AI%20algorithms%20analyze%20vast%20datasets,to%20more%20accurate%20weather%20forecasts>. Accessed 13 Dec. 2024.
- Buolamwini, Joy. *Unmasking AI: my mission to protect what is human in a world of machines*. Random House, 2023.
- Burnham, Kristin. "How artificial intelligence is transforming logistics" *MIT Sloan Management Review*, Climavision, 20 Aug. 2024,
<https://mitsloan.mit.edu/ideas-made-to-matter/how-artificial-intelligence-transforming-logistics>. Accessed 13 Dec. 2024.
- Sunday, Victor. "AI Navigation Apps Are Transforming Road Trips." *Mapmetrics.org*, 31 Mar. 2024,
<https://mapmetrics.org/blog/ai-navigation-apps-are-transforming-road-trips/>. Accessed 13 Dec. 2024.