Unmasking AI: My Mission to Protect What Is Human in a World of Machines
By Joy Buolamwini
Random House, 2023. 317 pages.

Unmasking AI: Why Ethical Tech Is Everyone's Responsibility

By Angela Nguyen

"With safety in sight and security calling, would you turn back for the forgotten ones?" (1). This opening question in Dr. Joy Buolamwini's *Unmasking AI* captures a crucial theme in her work—the responsibility to ensure no one is excluded from technological advancements. For undergraduates grappling with ethical technology, including myself, her work empowers us to confront the "coded gaze": "a reflection of the priorities, preferences, and prejudices of those who have the power to shape technology" (xii). As she tells undergraduate Agent Deb Raji, "Many people can gain tech skills. Your ability to communicate and tell a story about your technical work is what will separate you from your peers. It's what makes people see why change is necessary" (184). Her book reminds us that the stories we tell about technical work can bridge innovation and social impact.

Ensuring the representation of all groups requires everyone's input, as well as interdisciplinary expertise—not just technologists. One story recounts how Dr. Buolamwini helps tenants in Brooklyn understand the potential for algorithmic discrimination in an apartment complex populated mostly by people of color, women, and the elderly. Marginalized communities often lack awareness of algorithmic risks, and I admire Dr. Buolamwini's solidarity with them. She addresses issues such as the "model minority myth," discrimination against low-income Asians, higher SAT prep prices, false facial recognition matches, and advocates for people with disabilities, non-Western perspectives, and youth involvement.

My favorite part of the book is how Dr. Buolamwini shares her thought process—whether in designing the *Gender Shades* study to expose flaws in commercial facial recognition systems or connecting childhood anecdotes to key events. Her reflections deepen our understanding of the research process and its moral implications.

Her insights also extend to the complexity of human identity in relation to labeling and data: "Exploring labels around social constructs like race and ethnicity revealed the sociotechnical nature of the work I was doing" (113). This aligns with Audre Lorde's words, which she quotes: "People do not live single-issue lives." Technology, as Buolamwini demonstrates, is sociotechnical. It requires intention to prevent harm and is more complex than lines of code or binary answers. This is also why one need not be a computer scientist to contribute meaningfully to the field.

While tech companies race toward superintelligence, Dr. Buolamwini shows how current systems already fail marginalized groups, leading to false arrests, biased policing, and flawed medical diagnoses. I often ask myself: Why build on unsatisfactory foundations, only to fix them later? The profit-first mentality of these companies has led to short-term gains at the expense of long-term harm. Dr. Buolamwini emphasizes that only legal accountability has acted as a true deterrent. For example, Italian regulators paused ChatGPT over privacy concerns and fined Clearview AI for scraping photos of Italian residents. Meta was forced to delete more than a billion faceprints after Illinois enforced its

Biometric Information Privacy Act. And recently, FTC Chair Lina Khan has challenged big tech companies through antitrust cases for monopolistic practices—proof that it's still possible to hold them accountable.

Each regulatory success affirms the power of people from different backgrounds and disciplines to intervene in AI development, design, and deployment.

We must therefore keep the focus on the social impact of AI—not just the speculative futures some insist are already here. Dr. Buolamwini criticizes the effective altruism (EA) movement, which emphasizes long-term existential risks over addressing immediate harms. From my experience in both AI safety and tech ethics spaces, I believe the two can go hand in hand (what I'd call "sociotechnical safety")—but our starting point must be safety in design and all stages of the development cycle.

Dr. Buolamwini stresses that solving immediate issues is essential to addressing long-term ones. For example, she highlights lethal autonomous weapons as proof that existential risk comes not from hypothetical AI superintelligence, but from current human decisions. Military contracts with companies like Google, Amazon, and Microsoft already embody this danger. "Sadly, we do not need AI systems to have superintelligence for them to have fatal outcomes on individual lives" (150), she writes, underscoring the real, present-day harms of AI systems—many of which are being shaped and funded by universities like my own, Stanford.

Reading this book sharpened the dilemma many of us in tech face: how to work in systems that may perpetuate harm to our communities. Yet Dr. Buolamwini offers a powerful alternative—technology in the public interest. From conducting trachoma assessments in Ethiopia to presenting her findings at IBM and testifying before Congress, she shows that meaningful change is possible. Her question—"But in using the faces, was I any different than Facebook?"—asks us to confront the difference between coercive data extraction and informed, ethical research. While computer scientists freely scrape and monetize human face data under manipulative privacy policies, sociologists must navigate rigorous consent and oversight processes just to study people (102). The contrast is stark—and telling.

While this is a book review, it is also a call to action. As Dr. Buolamwini declares, "If you have a face, you have a place in the conversation [about AI]" (279). Technology impacts everyone—regardless of age, gender, race, ability, income, or background. It's our shared responsibility to shape a more inclusive technological future.

We must remain informed about the tech industry's most pressing issues. Decisions made by executives and policymakers today are already jeopardizing progress on diversity, inclusion, and combating algorithmic bias. These challenges transcend political divisions, making it essential for every community to contribute its voice to the future of technology.

To my fellow peers, *Unmasking AI* is not just a critique of tech's shortcomings—it's a roadmap for becoming part of the solution. When an undergraduate once asked how to get involved, Dr. Buolamwini's advice was clear: "Question the status quo, reach out and listen to people who experience algorithmic harms, check your assumptions, and explore intellectual terrain beyond computer science" (269).

Together, we can shape technology that prioritizes humanity over profit. As Dr. Buolamwini's own journey shows—from initially dismissing policy to becoming one of the leading voices in AI ethics—the future of AI is not fixed. It will reflect the values we choose to champion.