

Redistribution and Rekognition: A Feminist Critique of Algorithmic Fairness

Sarah Myers West

AI Now Institute at New York University

sarah@ainowinstitute.org

Abstract

Computer scientists, and artificial intelligence researchers in particular, have a predisposition for adopting precise, fixed definitions to serve as classifiers (Agre, 1997; Broussard, 2018). But classification is an enactment of power; it orders human interaction in ways that produce advantage and suffering (Bowker & Star, 1999). In so doing, it attempts to create order out of the messiness of human life, masking the work of the people involved in training machine learning systems, and hiding the uneven distribution of its impacts on communities (A. Taylor, 2018; Gray, 2019; Roberts, 2019). Feminist scholars, and particularly feminist scholars of color, have made powerful critiques of the ways in which artificial intelligence systems formalize, classify, and amplify historical forms of discrimination and act to reify and amplify existing forms of social inequality (Eubanks, 2017; Benjamin, 2019; Noble, 2018). In response, the machine learning community has begun to address claims of algorithmic bias under the rubric of fairness, accountability, and transparency. But it has dealt with these claims largely using computational approaches that obscure difference. Inequality is reflected and amplified in algorithmic systems in ways that exceed the capacity of statistical methods alone. This article examines how patterns of exclusion and erasure in algorithmic systems recapitulate and magnify a history of discrimination and erasure in the field of artificial intelligence, and in society more broadly. Shifting from individualized notions of fairness to more situated modeling of algorithmic remediation might create spaces of possibility for new forms of solidarity and

West, Sarah Myers (2020). Redistribution and Rekognition: A Feminist Critique of Algorithmic Fairness. *Catalyst: Feminism, Theory, Technoscience*, 6(2), 1–24.

<http://www.catalystjournal.org> | ISSN: 2380-3312

© Sarah Myers West, 2020 | Licensed to the Catalyst Project under a Creative Commons Attribution Non-Commercial No Derivatives license

refusal, accounting for the socially embedded nature of our experiences and seeking to build collective power across them rather than seeking to normalize and erase our differences.

Introduction

As artificial intelligence systems are deployed around us, they exhibit the same forms of inequality and discrimination that many of us experience in everyday life—only in ways that are more invasive, scale more rapidly, and are much less visible to us. What Safiya Umoja Noble (2018) calls *algorithmic oppression* exists in myriad forms, affecting the lives of women, trans people, people of color, and people with disabilities in both critical and mundane ways.

I recently compiled a list of examples:

- 2009: LA-based blogger Joz Wang (2009) wrote that the new Nikon S360 camera she'd bought her mom for Mother's Day kept asking her family members "Did someone blink?", even though their eyes were always open. Safiya Umoja Noble (2012) observed that her searches for the terms "Black girls" and "Latina girls" in Google's search engine turned up blatantly sexist and pornographic results.
- 2013: Latanya Sweeney (2013) found that Google searches for Black-identifying names were systematically more likely to retrieve ads suggestive of an arrest record.
- 2015: Jacky Alciné (2015) logged in to Google Photos to find a new set of albums tagged by the company's facial recognition software, including one that categorized him and his friend as "Gorillas."
- 2016: Microsoft's AI chatbot Tay learned to use racist and misogynist language in a single day (Vincent, 2016), and ProPublica published an investigative article reporting on the racially discriminatory risk assessment system COMPAS, which they found was twice as likely to falsely label Black defendants as future criminals than white defendants. The algorithm was used in decision-making at every stage in the criminal justice system, from setting bond amounts to evaluating parole (Angwin et al., 2016).
- 2017: Virginia Eubanks (2017) published *Automating Inequality*, which delved in detail into the automated decision-making systems used in the United States to police the poor, including systems automating eligibility for welfare benefits, housing, and child welfare.
- 2018: Timnit Gebru and Joy Buolamwini (2018) published the Gender Shades study showing that leading facial recognition software had higher failure rates to recognize women with darker skin pigmentation; Trans drivers logging in to Uber found they were locked out due to a new security feature that relied on facial recognition software that no longer identified them (Vincent, 2017); and a team of

researchers found systematic discrimination in Facebook's delivery of housing and employment ads, leading to skewed results that delivered ads for higher-paying jobs to white men and lower-paying jobs to women and people of color (Ali et al., 2018).

- 2019: Inioluwa Deborah Raji and Joy Buolamwini (2019) published an audit of Amazon's Rekognition facial analysis software, widely used by law enforcement agencies, finding it, too, had significant failures along gender and race. The same year, the company heavily marketed the tool to US Immigrations and Customs Enforcement to accelerate deportations and announced its "emotion detection" software could now detect fear (Rose, 2019).

These examples are only a selection of an exponentially growing list of forms of algorithmic oppression. Even then, they represent only those that have surfaced in systems long after they've been deployed, visible enough to attract public attention.

As examples of algorithmic oppression are rapidly revealed around us, it is clear that their effects are not evenly distributed. Across them, women and gender minorities, people of color, people of lower socioeconomic status, and people with disabilities bear the burden of algorithmic exclusion (Gandy, 1993)—and, as elsewhere, these burdens are particularly acute for those who lie at the intersections of these forms of identity (Crenshaw, 1991). I argued in (West et al., 2019) that there is a feedback loop between the notably non-inclusive and actively toxic workplaces in which many of these systems are being built and the algorithmic systems that reproduce manifold forms of oppression and discrimination. In this article, I'd like to trouble that notion, digging deeper into the messiness of how these elements are both co-constituted and a function of the dynamics of capitalism: the patterns of exclusion and erasure we see in algorithmic systems recapitulate and magnify a history of discrimination and erasure in the field of artificial intelligence, and in society more broadly.

In examining this history, I was initially interested in understanding how historic gender-based discrimination in AI workplaces may have shaped AI systems' literal failures to "see" and "hear" women. But as I dug deeper, reading the work of scholars who had long been critiquing technological discrimination, I recognized how much my understanding of algorithmic oppression was limited by how many systems worked *well* for me as a white, cisgendered woman based in the United States—and that making AI systems "work better" wasn't necessarily the right objective to be seeking at all.¹

As scholars like Noble and others have noted, the kinds of harms enacted by algorithmic oppression require much more than making them unbiased, or inclusive—more than “fixing” broken systems or seeking to achieve predefined metrics of algorithmic fairness (a term of art used in the machine learning community to describe research and applications that attempt to address discrimination in algorithmic systems). Reading across critiques of algorithmic fairness (eg., Barocas & Selbst, 2016; Noble, 2018; Hoffmann, 2019; Benjamin, 2019), my aim in this piece is to contextualize one possible path toward redressing algorithmic oppression in a way that accounts for and acknowledges the role of power in shaping the differential experiences we each have with algorithmic systems.

A situated approach that draws on the work of feminist STS scholars may offer a path toward a more comprehensive way to address the harms enacted by algorithmic systems. Many current approaches focus on fairness through achieving statistical equity. I argue that in place of fairness, remediation—the work of making things right, of stopping and reversing the damage done by harmful AI systems—offers a powerful alternative. A path toward remediation must account for the ways in which recognition is not evenly distributed—that what may be “fair” in one context is may be deeply unfair in others.² A situated model of knowledge production necessitates a more complicated mode of remediation than fairness offers; it requires the acknowledgement of the highly contextual and entangled nature of our experiences with algorithmic systems, and the ways in which they are deeply intertwined with all forms of power.

Algorithmic Fairness and Its Limitations

Computer scientists, and artificial intelligence researchers in particular, have a predisposition for adopting precise, fixed definitions to serve as classifiers (Agre, 1997; Broussard, 2018). But classification is an enactment of power; it structures human interaction in ways that produce advantage and suffering (Bowker & Star, 1999). Classification aims to create order out of the messiness of human life, and in so doing, hides the uneven distribution of its impacts on communities. The intricate processes of defining features, preclassifying training data, and adjusting thresholds and parameters involve many kinds of human judgments at multiple scales (Burrell, 2016), incorporating not only the design choices of engineers but micro-level choices by those engaged in the work of training and calibrating machine learning systems (Barocas & Selbst, 2016; A. Taylor, 2018; Gray & Suri, 2019; Roberts, 2019). Put simply, classification is foundationally a social process that is entwined with power at every level.

Addressing biases introduced through the work of classification is a primary focus of research on fairness, accountability, and transparency in artificial intelligence systems. In 2014 a collective of machine learning researchers convened at a workshop at the Neural Information Processing Symposium, one of the top conferences in the field, to grapple with the anxieties surrounding the use of machine learning in decision-making processes in core social domains.

Anticipating an explosion of challenges in the years to come, a provocation outlined the motivations behind the workshop: decisions informed by what they called “big data” could create discriminatory harms—even if unintended—that could result in less favorable treatment for already disadvantaged groups. Sited at one of the top academic conferences in the field of machine learning, the workshop foregrounded computational approaches to social problems. It aimed to focus on “challenging constraints on the practical application of machine learning” and “problems that can lend themselves to novel computational solutions” (Hardt & Barocas, 2014).

Importantly, *Fairness, Accountability and Transparency in Machine Learning (FAT/ML)* took on an explicitly interdisciplinary frame, curating a collective of speakers from a mixture of technical and legal backgrounds. The work initiated at FAT/ML migrated across a series of different venues in the years following—what became an annual workshop at NeurIPS, a conference on neural information processing, was soon accompanied by subsequent workshops on recommender systems and data science, as well as an Association for Computing Machinery (ACM)-sponsored conference, FAccT (previously known as FAT*).

Although discourses around algorithmic “fairness” are increasingly interdisciplinary, a dominant current has centered on computational framings of the problem, such as the development of tools designed to identify and mitigate bias in algorithmic systems. These computational approaches to fairness aim, for example, to produce algorithmic systems that perform an agreed-on level of parity, through statistical modeling or by correcting statistical bias in datasets so that accuracy rates are achieved across demographic categories (Dwork et al., 2012). Anna Lauren Hoffmann notes a number of limitations in the model of anti-discrimination adopted in these discussions, observing that efforts to combat algorithmic oppression “fail to address the very hierarchical logic that produced advantaged and disadvantaged subjects in the first place” (2019, p. 901). Computational definitions of fairness often box out deeply significant social and structural concerns, weakening the underlying meaning of fairness.

These limitations are in many respects a legacy of fairness tests that developed as a proxy for proving discrimination in the wake of the Civil Rights Act of 1964 (Hutchinson & Mitchell, 2018; Hoffmann, 2019). Ambiguities around the legal standards adopted in education, hiring, and housing provoked the development of a market for these quantitative tests as proxies for discrimination against members of protected classes—a simplification of much deeper systemic racial and gender inequalities. The adoption of these standards spurred a body of research—and before long, an entire industry—focused on measuring “unfairness” in educational and employment testing, and “disparate impact” in the realms of employment and housing (Freeman, 1978; Hutchinson & Mitchell, 2018). We similarly see an industry emerging that promises fairness measurement and recalibration, through products like IBM’s AI Fairness 360 Toolkit, Accenture’s Teach and Test services, and Facebook’s Fairness Flow tool.

Fairness tests in education, hiring, and employment came about, at least in part, because proving discrimination under current US civil rights law is so hard to do. But rather than remediate the underlying legal and social structures perpetuating inequality, the fairness measures used to prove statistical disparities under anti-discrimination laws have instead been slowly gutted—ultimately rendering US civil rights law a limited mechanism for redressing algorithmic harms that impact communities and individuals far beyond the United States (Richardson, 2020; Barocas & Selbst, 2016).

Moreover, once a model is deemed “fair,” it may still be used to enact other kinds of harms, such as by being incorporated into systems for the policing of communities that already experience discrimination. Plug-and-play approaches to “fairness” distract from the longer, slower, harder work of grappling with the systemic and structural dimensions of discrimination. As Ruha Benjamin observes, “Computational approaches to a wide array of problems are seen as not only good but necessary, and a key feature of cost-cutting measures is the outsourcing of decisions to ‘smart’ machines...Profit maximization, in short, is rebranded as bias minimization” (2019, p. 30). In them, inequality is reflected, amplified, and enacted in algorithmic systems in ways that statistical methods can only partially address.

These simplified approaches to fairness are undesirable for other reasons. They don’t adequately match the actual conditions of developing and implementing algorithmic systems. The streamlined models of data training, model

construction, model testing and bias mitigation, system deployment, and feedback do not account for the incredibly complex set of actors and activities involved in the process of doing data science (Holstein et al., 2018; Passi & Barocas, 2019).

They don't account for the harms that surface not in data, but in the design of the system itself. They enact a kind of violence by actively reproducing ideas that damage vulnerable communities, or by preventing people from meeting their basic needs (O'Neil, 2016; Onuoha, 2017; Hoffmann, 2018). For example, automated gender recognition systems encode the notion that gender is a binary, immutable, physiological form of identity, an erasure of the experiences of trans people with very real and systematic consequences (Keyes, 2018).

They also don't account for the ways that attempting to fix bias, such as producing more diverse datasets, can itself create inequality (Hoffmann, 2020). Ben Green (2020) describes this as *the impossibility of fairness*—that, in an unequal society, decisions rooted in formal equality will produce substantive inequality. This is exemplified in the problem of diversity in datasets: building datasets is expensive and time-consuming, and engineers are incentivized to utilize the data already available to them. This has led to the use of harmful practices in order to produce the large and diverse data sets needed to train machine learning models—for example, by incentivizing more invasive surveillance of the communities already discriminated against by an algorithmic system, or by scraping their data without their consent. For example, one research team sought to resolve the failure of a facial recognition system to identify the faces of trans individuals by scraping YouTube videos posted by users undergoing the transitioning process (Vincent, 2017). This not only perpetuates existing forms of inequality, but can lead to re-enacting harms on the communities that are already discriminated against by algorithmic systems.

Finally, fairness discourses insufficiently account for the ways that breakdowns in AI systems can be productive. As scholars like Alondra Nelson, Joy Buolamwini, and Timnit Gebru have observed, the objective is not to make “fairer” or “more inclusive” systems of surveillance for policing communities of color. Breakages can afford a new kind of power, to evade the gaze of pervasive surveillance that exists in both cameras and in code. Patricia Hill Collins (1986) reminds us that the act of refusal in favor of self-definition is itself a profoundly political act. In tracing through the outlines of Black feminist thought, she emphasizes the meaning of the rejection of external definitions of Black womanhood:

the individual women who in their consciousness choose to be self-defined and self-evaluating are, in fact, activists. They are retaining a grip over their definition as subjects, as full humans, and rejecting definitions of themselves as the objectified “other”...People who view themselves as fully human, as subjects, become activists, no matter how limited the sphere of their activism may be. (Collins, 1986, p. S24)

We may similarly create spaces for self-definition through the rejection of classification by artificial intelligence systems.

Yet, in other times and places, the breakdown of systems can put people at risk and cause harm by denying the standing of entire groups of people (Young, 1990), or reflecting to them a confining or demeaning picture of themselves (C. Taylor, 1992). For example, trans scholars such as Sasha Costanza-Chock (2018) document how the TSA’s millimeter wave scanning system marks trans bodies as more risky—and thus subject to more invasive policing—than other kinds of bodies. And in other cases still, the very *existence* of certain kinds of AI systems enacts a kind of violence: Os Keyes (2018) and Hamidi et al. (2018) each make this observation about the harms enacted by automated gender recognition systems that erase trans people. Algorithmic systems can thus play a legitimating, discursive, and dignitary function in which recognition is deeply meaningful (Hoffmann, 2018; Dencik et al., 2018).

For these reasons, many scholars whose work is centrally concerned with data justice are concluding that fairness discourses are a limiting frame for grappling with the very real harms enacted by artificial intelligence systems (Noble, 2018; Benjamin, 2019; Hoffmann, 2019). They leverage critiques by thinkers such as Iris Marion Young (1990) to argue that a conception of justice must begin with the concepts of domination and oppression, rather than the current focus on the *distribution* of harms, or more rarely, advantages. Thus, the path forward necessitates situating discrimination in artificial intelligence systems in context—beginning with a better understanding of the environments in which these systems are themselves developed.

Discrimination in the History of AI

Artificial intelligence systems do not only reflect and amplify the inequalities of today (Barocas & Selbst, 2016); gendered, racialized, and classed forms of discrimination are built into the field of artificial intelligence at a foundational level. In this analysis, I focus narrowly on forms of gender-based discrimination in

the history of AI, however this focus should not discount the influence of race, disability, and other forms of identity-based discrimination in shaping the field.³

Algorithmic oppressions can be traced across space and time, from early voice recognition systems calibrated such that women's voices were often, quite literally, unheard, to associations between early face recognition systems and principles of physiognomy (Gates, 2011). Where women *are* recognized by AI systems, they often are situated in positions with relatively little power, such as the overwhelming use of female characters for virtual assistants since the early 2000s in ways that replicate gendered stereotypes about women in positions of servitude (Zdenek, 2007; Steele, 2018; Sweeney, 2018; UNESCO, 2019).

This mirrors a practiced erasure of women from the field of AI over the course of its decades-long history. From its earliest inception, artificial intelligence research culture was male-dominated: the relationships between university-based computer science labs and the institutions of military service that funded them created mutually reinforcing cultures of masculinity (Edwards, 1990). The field's original conception as "automata studies" itself draws on highly gendered constructs: Silvia Federici (2004) observes that the emergence of the concept of automata in the sixteenth century relied on notions of scientific rationalization developed by Hobbes and Descartes. In the pursuit of modeling rational behavior through machines, early scientists borrowed from a particular conception of human intelligence grounded in white, upper-class, adult masculinity.

Centuries later, the models of cognitive rationalism that shaped early AI research on expert systems drew on Cartesian models of intelligence that sought to replicate human thought through formal mathematical representations (Edwards, 1990). Though in its early years the field was remarkably interdisciplinary, borrowing from psychology, philosophy, linguistics, and neuroscience in addition to mathematics and computing, from at least the mid-twentieth century most of the figureheads of artificial intelligence were primarily concerned with developing modes of mathematical representation of the human mind that could be represented through programmable logics.

Though the field has undergone many shifts over the years, it has largely retained this technical and rationalized orientation: in her study of the history of computational text prediction and automatic speech recognition, Xiaochang Li (2018) finds a "statistical turn" in the field, during which there was a conscious effort to exclude other disciplines when collaborations proved slower and more

arduous, producing a radical data orthodoxy among AI researchers. Advanced technical degrees are now central to credentialing for the field, even as AI researchers engage in building tools that have implications for other domains like law, sociology, and medicine.

The technical orientation of the field of artificial intelligence has social implications far beyond the now-popular notion that more women and people of color need to “learn to code.” This rhetoric positions the absence of diversity as the product of disinterest or lack of skills acquisition on the part of “diverse” communities, seeking to address it through programming designed to provide these skills (rather than address other, deeper structural factors that are as influential) (West et al., 2019). But we already know that the “technical” is an active construction, built around practices of exclusion honed over centuries. Feminist STS scholars have shown how technical knowledge is socially constructed as an inherently masculine domain, and justified under rubrics such as meritocracy (Wacjman, 1991; Oldenziel, 1997; Cockburn, 2009). There is nothing inherent about the association between masculinity and technical skill. Rather, the definition of a domain as “technical” in fact *follows from* discrimination: domains that are ascribed as primarily masculine are defined as higher-skilled and more technical than those seen as feminine, regardless of their actual content (Cowan, 1976; Cockburn, 1985; Durack, 1997/2009; Hicks, 2017). Skill is an ideological category imposed on types of work by the identities of the workers who perform it (Phillips & Taylor, 1980). As Mar Hicks writes in their book *Programmed Inequality*, “throughout history, it has often not been the content of the work but the identity of the worker performing it that determined its status” (2017, p. 16).

In the history of the field of computer programming, there was an observable shift in the status of software development when the increasing demand for software systems led to a new appreciation for programmers and rise in salaries, sparking a transition from a field largely populated by women to one that predominantly employed men (Light, 1999; Ensmenger, 2010, 2015). Women began to be portrayed as low-skilled and low-wage employees alongside the development of new cultural tropes of “geek masculinity,” bolstering the status of male computer nerds, in which individual artistic genius, eccentricity, anti-authoritarian behavior, and dislike of activities involving human interaction acted to bolster male software programmers’ personal and professional authority. These masculinist histories, largely originating from Silicon Valley, obscured vibrant computer cultures elsewhere in which women played a central role (Rankin, 2018).

Similar processes of “deskilling” are mirrored in the current state of the AI field, which largely excludes the labor of overwhelmingly woman-identifying contract workers who perform work critical to labeling and preparing the data upon which machine learning systems are trained (Gray, 2015; Irani, 2015; Roberts, 2016). Astra Taylor (2018) recently coined the term “fauxtimation” to describe the erasure of this labor, aptly tying it to the image of the Mechanical Turk, a robot that purported to know how to play chess, but was in fact a hoax designed to mask the human player hidden below its board. Importantly, she makes a bridge between the erasure of women in the AI field and capitalist practice: as she puts it, fauxtimation “reinforces the perception that work is unpaid and acclimates us to the idea that one day we won’t be needed” (A. Taylor, 2018), speeding up a process that pushes gendered, racialized, and classed bodies to the margins of society.

Patterns of Erasure in AI

In the 1980s and 1990s women began to organize and raise the visibility of discrimination in the labs where early AI systems were being constructed. For example, female graduate students in the computer science and artificial intelligence labs at MIT meticulously documented their experiences with discrimination in the 1983 report, “Barriers to Equality in Academia: Women in Computer Science at MIT.” In the report, the students remarked that “Many women are treated as if they were invisible in technical situations...[O]verlooked in technical discussions and excluded from group efforts, their work is attributed to male colleagues, and their opinions are not sought on relevant technical subjects.” (1983, p. 9) Moreover, they felt that women in the department were caught in a double bind; they were expected by some to adopt a “masculine” style to achieve credibility, while others felt that this demeanor would be perceived as “aggressive” because it did not conform to feminine expectations (MIT, 1983).

Ellen Spertus’s (1991) “Why Are There so Few Women in Computer Science?” further depicts barriers to gender diversity in the field. Though it addresses computer science in general, it was published under the imprimatur of MIT’s Artificial Intelligence Technical Reports, and its data provides sporadic insight into the state of AI in the late 1980s and early 1990s: in a section on the impact of gendered language, Spertus said that over one month of observation, she observed many examples of people at the MIT AI Lab attempting to use gender-neutral language in place of masculine defaults, but failing to do so in a consistent way. A female computer scientist whom Spertus interviewed noted that none of the major conferences in the field of AI offered childcare, making it difficult for her

to participate in professionalization opportunities, while another referenced walking out after a sexist aside was made by a guest lecturer in MIT's core AI class.

In 1992 an editor at *IEEE Expert* led a study of gender diversity in the field of AI, and found that only 13 percent of published authors in the journal over the previous four years were women (Strok, 1992). Moreover, he found that in 1992, women represented

- 10 percent of program committee members at twenty-five AI-related conferences over the previous three years;
- 8 percent of invited speakers at seven AI-related conferences;
- 9 percent of the editorial board members of seventeen AI journals and magazines;
- 9 percent of authors in the leading journals *IEEE Expert*, *Artificial Intelligence*, and *AI Magazine* over four years;
- 25 percent of Association for the Advancement of Artificial Intelligence (AAAI) executive councilors; and
- 8 percent of AAAI fellows. (Strok, 1992)

The context surrounding the *IEEE* survey provides some insight into the kinds of challenges experienced by women in the field at the time: Derek Strok, the survey's author and an editor at *IEEE Expert*, was motivated to examine the issue after a woman whose article had been accepted to the journal asked for her initials rather than her full name to be used so that readers would not question the validity of her work. Taking an informal survey of contributors to the journal, he reported that "dozens" of women were willing to discuss their interests and achievements, but many did not want to be quoted on the record about gender-based obstacles for fear of losing their jobs. The women said that barriers to their inclusion were rampant: women reported frequently encountering discrimination in the workplace, regular displays of sexist material such as the use of lewd imagery in coworkers' Windows screensaver displays, and the frequent employment of gender stereotypes in the examples for course material. Often, systems of male patronage were seen as a requirement for women to "make it" in the field; those who succeeded had prominent advisors in the original inner circle of researchers at institutions such as MIT, Stanford, and Carnegie Mellon University.

Despite these clear obstacles, in his conclusion Strok placed the onus for solving the diversity problem primarily on women: "The underrepresentation problem does not derive from the fact that female scientists are consciously ignored, but that (1) unconscious biases drive some women to quit science, and (2) not enough

women (and minorities) are entering scientific fields in the first place," he concluded, issuing recommendations from the women interviewed such as "be good at what you do," "find a support network," and "get good training" (Strok 1992, pp. 16-21). These solutions seem paltry in view of the evidence of structural discrimination presented elsewhere in the article: sexism, fear of retaliation, and widespread barriers to promotion seem as likely to be the cause of the limited representation of women in the field.

Perhaps the most thorough observations of what she called the "deletion of women" in AI are outlined in Diana Forsythe's (2002) ethnographic work at five bioinformatics labs over the course of the 1990s. This "deletion" took many forms: Forsythe remarked that the work of women administrators, critical to the grant-seeking process that kept the labs alive, was deemed so unimportant that one researcher remarked a lab that had multiple women serving in an administrative capacity had no women employees at all. She also cites multiple instances of harassment and discrimination in the laboratory environment: women grad students were targeted by male students with sexualized humor, such as the installation of sexualized sound files on their computer that could not be turned off once clicked on. Male grad students also installed screensavers with lewd images of women. When the behavior was escalated to supervisors, they often wrote it off as harmless or a practical joke.

Though these accounts are in themselves concerning, it is Forsythe's documentation of the less visible forms of discrimination experienced in AI labs that is most illuminating. For example, she found that the PhD degrees held predominantly by women in the lab were discounted or seen as "less serious" than the MD degrees held predominantly by men (Forsythe, 2002). Moreover, women in the lab were often defined in terms that related them to other men, while men were defined by what they do. These systems of patronage had profound effects on hiring patterns, as Forsythe observed that mentorship actually hindered women from promotion. She saw a reluctance to hire women away from other labs as though their (male) supervisors had "ownership" over them. Forsythe's findings are consonant with longitudinal studies that have found that though women are more likely than men to have mentors, men who receive mentorship are statistically more likely to be promoted, while women who receive mentorship are in lower-level positions, paid less, less satisfied, and less likely to be promoted than men (Ibarra et al., 2010). Forsythe also found that women researchers often experienced pressure to pursue their mentor's interests rather than their own, and to include men as authors on papers on which they did not do work—practices

that could significantly inhibit the likelihood of women researchers' work being acknowledged by others and impact their promotion.

Though each of these studies provides a distinct, though partial, view of the state of diversity in artificial intelligence, they agree that gender discrimination and harassment are significant challenges that have plagued the field of AI over the course of decades. Moreover, these studies point to several specific structural barriers that combined to prevent the advancement of women in the field: overt harassment, a lack of resources, constraints on the visibility of female researchers' accomplishments and limits on their opportunities for promotion, among others. They are, admittedly, largely informal and out of date—but this is itself reflective of the exclusion of gender from the central concerns of the field. Controversies over the treatment of women in the field of AI stretch back decades, and yet even as issues of bias, fairness, accountability, and transparency became central objects of inquiry, we still lack substantive research that examines these patterns of discrimination, or that examines how AI systems are experienced differently on the basis of gender, race, class, and sexuality.

It is worth considering how re-examining AI to account for this exclusion may prompt new ways of thinking about algorithmic oppression. Exclusions have deeply shaped what we understand to be artificial intelligence, as they shaped the field's conceptions of what constitutes intelligence itself. As feminist scholars have long pointed out, the ideas about universality, transparency, and neutrality that ground cognitive rationalism reinforce a patriarchal view of science: that what we understand to be rational or objective thought has worked historically to diminish women's contributions to the field and to exclude domains typically coded as women's work from science and technology. The definitions of intelligence dominant in AI are premised on deeply gendered and racialized constructs, resulting in a flattened approach to resolving the challenge posed by algorithmic oppression.

Recognition and Remediation

In the 1990s, scholars such as Nancy Fraser, Axel Honneth, and Iris Marion Young engaged in a debate around the question of "redistribution or recognition": across a series of essays, they questioned whether the pursuit of representation and identity politics served, or supplanted, the underlying cause of seeking justice through redistribution of resources. This debate highlighted that neither can be separated from one another: redistribution and recognition are endlessly entangled, and neither alone is sufficient to achieve justice. Distribution alone

insufficiently accounts for the cause of injustice (Young, 1990, 1997; Fraser, 1995, 1997; Honneth, 1995; Fraser & Honneth, 2003).

A similar tension is palpable in the discussions around algorithmic oppression in the distinctions drawn, for example, between allocative and representational harms (Barocas et al., 2017). Most conceptualizations of fairness outlined above adopt only the logic of redistribution, by seeking to account for and equalize across differences in the harms enacted by AI systems or the benefits allocated by them. As such, they cannot mitigate the effects of the larger structures in which AI systems are deployed, such as increased surveillance and policing of the same communities underrepresented in AI systems; nor can they account for the impact of erasures among those involved in the construction of AI systems. A more comprehensive accounting of algorithmic harms must thus grapple with the ways in which the logics of recognition are *themselves* not evenly distributed.

A situated model of knowledge production can provide a more robust way to hold the institutions creating and deploying algorithmic systems to account by affirming difference, rather than papering it over. This necessitates a more complicated approach to fairness that acknowledges the highly contextual and entangled nature of our experiences with algorithmic systems, and the ways in which they are deeply intertwined with power. And though many within the fairness community are beginning to explore things along these lines, their work perhaps looks less like *fairness* and more like *remediation*—the work of recovery, recompense, stopping, and reversing the harms caused by algorithmic systems rather than seeking to repair them.

What would remediation look like in practice? A starting place that some scholars have proposed is to redefine artificial intelligence to encompass the wealth of experience that those who have largely been erased from the history of artificial intelligence can bring: women, trans people, people of color, the disability community. Remediating this exclusion may facilitate a more contingent and collective notion of intelligence. Inspired by feminist standpoint theory (Hartsock, 1983; Collins, 1986), which argues that the social order looks different depending on our position within it, scholars such as Donna Haraway have long advocated for new ways of thinking about knowledge that draw upon each of our limited locations and standpoints to provide more adequate, sustained, and “objective” accounts of the world. As Haraway puts it, “feminists have stakes in a successor science project that offers a more adequate, richer, better account of a world, in order to live in it well and in critical, reflexive relation to our own as well as others’

practices of domination and the unequal parts of privilege and oppression that make up all positions" (1988, p. 579).

She argues for the notion of *situated knowledges*, learning together by collectively joining our partial views of the world into a larger whole. As she puts it, "rational knowledge does not pretend to disengagement: to be from everywhere and so nowhere, to be free from interpretation, from being represented, to be fully self-contained or fully formalizable. Rational knowledge is a process of ongoing critical interpretation among 'fields' of interpreters and decoders" (Haraway, 1988, p. 590). Lucy Suchman makes a similar argument, drawing on Haraway to develop a set of feminist principles for technology design using the notion of "located accountabilities." As Suchman puts it, "the only possible route to objectivity...is through collective knowledge of the specific locations of our respective visions. By extension, the only possibility for the creation of effective objects is through collective knowledge of the particular and multiple locations of their production and use" (2003, p. 96).

Only through joining each of our partial views into a collective subject position can we gain meaningful insight into the world around us. And in fact, this is attuned to the underlying logics of most of the AI systems we use today: our data is entangled with the data of those around us, across space and time. We rarely engage in simple, one-to-one interactions with artificial intelligence systems. Rather, we engage collectively, if unknowingly, with all those whose data shaped and trained them. Shifting from individualized notions of fairness to more situated modeling of algorithmic remediation might create spaces of possibility for new forms of solidarity and refusal, accounting for the socially embedded nature of our experiences and seeking to build collective power rather than seeking to normalize and erase our differences.

This work is already underway, among researchers who are answering Suchman's call to move beyond scholarly observation by getting inside the black boxes of technology production, to disrupt the engine rooms of technological production. In 2019 feminist researchers issued a Feminist Data Manifest-No, a collective declaration of refusal to engage with harmful data regimes and commitment to building new data futures (Cifor et al., 2019). A growing number of researchers of algorithmic systems are making calls for the development of frameworks to understand data science as a situated practice, including reorienting their research questions "upwards," away from those who experience systematic disadvantage and toward those who inhabit positions of power and authority

(Elish & Boyd, 2018; Barabas, 2020; Katell et al., 2020). Others call upon technical designers to redraw their abstraction boundaries to include social actors, and not just technical actors (Selbst et al., 2018), and to embed a decolonial critical approach within technical practice (Mohamed et al., 2020). Such work is critical to overcoming the deep structural barriers that inhibit the development of a feminist “AI from below.”

Re-imagining artificial intelligence to account for its exclusions will take more than fairness alone can offer. Bringing the projects of institutional accountability for inequities in tech workplaces, and algorithmic accountability addressing discrimination in technical systems, in conversation with one another under a rubric of remediation can enable more fundamental shifts in underlying power structures—shifts that allow us to address more deeply and fully how power is at work within the industry, and how it might be distributed to serve more just and equitable ends.

Acknowledgements

The author would like to thank Joy Rankin, Rashida Richardson and the anonymous reviewers for their feedback on the draft manuscript.

Notes

¹ It is also important to acknowledge that, in many respects, I am the beneficiary of a concerted effort to increase the number of “Women in Tech”—efforts that privilege US-based, white, cis, and able-bodied women like me over others. I thus engage my work with an explicit acknowledgement and commitment to not only feminist, but antiracist and anti-ableist principles, and in solidarity with those who experience the intersections of algorithmic oppressions more acutely than I do.

¹ See, for example, JafariNaimi et al., 2015.

¹ See, for example, Guillory, D. (2020) Combating Anti-Blackness in the AI Community. *Computers and Society*. <https://arxiv.org/abs/2006.16879>.

References

Alcine, J. (2015). Twitter. <https://twitter.com/jackyalcine/status/615329515909156865>

Ali, M., Sapiezynski, P., Bogen, M., Korolova, A., Mislove, A., & Rieke A. (2018). Discrimination through optimization: How Facebook’s ad delivery can lead to skewed

outcomes. <https://arxiv.org/pdf/1904.02095.pdf>

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias. *ProPublica*. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

Agre, P. (1997). *Computation and human experience*. Cambridge University Press.

Barabas, C. (2020). Studying up: Reorienting the study of algorithmic fairness around issues of power. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 167-176. <https://dl.acm.org/doi/abs/10.1145/3351095.3372859>

Barocas, S., & Selbst, A. (2016). Big Data's disparate impact. *California Law Review*, 671. <http://dx.doi.org/10.2139/ssrn.2477899>

Barocas, S., Crawford, K., Shapiro, A., & Wallach, A. (2017, Mar. 18-19). *The problem with bias: Allocative versus representational harms in machine learning*. [Paper presentation]. SIGCIS, Philadelphia, PA, United States. <http://meetings.sigcis.org/uploads/6/3/6/8/6368912/program.pdf>

Benjamin, R. (2019). *Race after technology: Abolitionist tools for the new Jim code*. Wiley Books.

Broussard, M. (2018). *Artificial unintelligence: How computers understand the world*. MIT Press.

Bowker, G., & Star, S. L. (1999). *Sorting things out: Classification and its consequences*. MIT Press.

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research*, 81, 1–15. <http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf>

Burrell, J. (2016). How the machine “thinks”: Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1): 1-12. DOI: 10.1177/2053951715622512

Cifor, M., Garcia, P., Cowan, T. L., Rault, J., Sutherland, T., Chan, A., Rode, J., Hoffmann, A.L., Salehi, N., & Nakamura, L. (2019). *Feminist data manifest-no*. <https://www.manifestno.com/>

Cockburn, C. (1985). *Machinery of dominance: Women, men, and technical know-how*. Boston: Northeastern Press.

Cockburn, C. (2009). On the machinery of dominance: Women, men, and technical know-how. *Women's Studies Quarterly*, 37(1&2), 269–273.

Collins, P. H. (1986). Learning from the outsider within: The sociological significance of Black feminist thought. *Social Problems*, 33(6), S14–S32.

Costanza-Chock, S. (2018, July 27). Design justice, AI, and escape from the matrix of domination. *Journal of Science and Design*.

<https://jods.mitpress.mit.edu/pub/costanza-chock>

Cowan, R. S. (1976). The "Industrial Revolution" in the home: Household technology and social change in the 20th century. *Technology and Culture*, 17(1), 1–23.

Crenshaw, K. (1991). Mapping the margins: Intersectionality, identity politics, and violence against women of color. *Stanford Law Review*, 43(6), 1241–1299

Dencik, L., Jansen, F., & Metcalfe, P. (2018, August 30). A conceptual framework for approaching social justice in an age of datafication. DATAJUSTICE project.

<https://datajusticeproject.net/2018/08/30/a-conceptual-framework-for-approaching-social-justice-in-an-age-of-datafication/>

Durack, K. T. (2009). Gender, technology, and the history of technical communication. *Technical Communication Quarterly*, 6(3), 249–260. (Original work published in 1997).

Dwork, C., Hardt, M., Pitassi, T., Reingold, O., & Zemel, R. (2012). Fairness through awareness. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference* (pp. 214–226). ACM.

Edwards, P. (1990). The army and the microworld: Computers and the politics of gender identity. *Signs*, 16(1), 102–127.

Elish, M. C., & boyd, d. (2018). Situating methods in the magic of big data and AI. *Communication Monographs*, 85(1), 57–80.

Ensmenger, N. (2010). *The computer boys take over: Computers, programmers, and the politics of technical expertise*. MIT Press.

Ensmenger, N. (2015). Beards, sandals, and other signs of rugged individualism: Masculine culture within the computing professions. *Osiris*, 30(1), 38–65.

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. New York: Macmillan Press.

Gray, M. L. (2015, November 12). The paradox of automation's "last mile." *Social Media Collective*. <https://socialmediacollective.org/2015/11/12/the-paradox-of-automations-last-mile/>

Gray, M., & Suri, S. (2019). *Ghost work: How to stop Silicon Valley from building a new global underclass*. Houghton Mifflin Harcourt.

Guillory, D. (2020) Combating anti-Blackness in the AI community. *Computers and Society*. <https://arxiv.org/abs/2006.16879>.

Federici, S. (2004). *Caliban and the witch*. Autonomedia.

Forsythe, D. (2002) *Studying those who study us: An anthropologist in the world of artificial intelligence*. Stanford University Press.

Fraser, N. (1995). From redistribution to recognition? Dilemmas of justice in a 'post-socialist' age. *New Left Review*, 212.

Fraser, N. (1997). A rejoinder to Iris Marion Young. *New Left Review*, 223.

Fraser, N., & Honneth, A. (2003). *Redistribution or recognition?: A political-philosophical exchange*. Verso Books.

Freeman, A. D. (1978). Legitimizing racial discrimination through antidiscrimination law: A critical review of Supreme Court doctrine. *Minnesota Law Review*, 62, 1049–1120.

Gandy, O. H., Jr. (1993). *The panoptic sort: A political economy of personal information. Critical studies in communication and in the cultural industries*. Westview Press.

Gates, K. (2011). *Our biometric future: Facial recognition technology and the culture of surveillance*. New York: NYU Press.

Green, B. (2020, January 27–30). *The false promise of risk assessments: Epistemic reform and the limits of fairness* [Paper presentation]. FAT* '20: Proceedings of the 2020 Conference on fairness, accountability, and transparency, Barcelona, Spain. <https://dl.acm.org/doi/abs/10.1145/3351095.3372869>.

Hamidi, F., Scheuerman, M.K., & Branham, S.M. (2018, Apr. 21-26). *Gender recognition or gender reductionism? The social implications of automatic gender recognition systems*. [Paper presentation]. CHI 2018, Montreal, CA. <https://dl.acm.org/doi/10.1145/3173574.3173582>

Haraway, D. (1988). Situated knowledges: The science question in feminism and the privilege of partial perspective. *Feminist Studies*, 14(3), 575-599.

Hardt, M. and Barocas, S. (2014, Dec. 12). *Fairness, Accountability, and Transparency in Machine Learning*. [Workshop]. Neural information processing symposium, Montreal, Canada.

Hartsock, N. C. M. (1983). The feminist standpoint: Developing the ground for a specifically feminist historical materialism. In M. B. Hintikka & S. Harding (Eds.), *Discovering reality: Feminist perspectives on epistemology, metaphysics, methodology, and Philosophy of Science*. (pp. 283–310). Springer.

Hicks, M. (2017). *Programmed Inequality: How Britain discarded women technologists and lost its edge in computing*. MIT Press.

Hoffmann, A. L. (2018, April 30). Data violence and how bad engineering choices can damage society. *Medium*. <https://medium.com/s/story/data-violence-and-how-bad-engineering-choices-can-damage-society-39e44150e1d4>

Hoffmann, A. L. (2019). Where fairness fails: Data, algorithms, and the limits of antidiscrimination discourse. *Information, Communication & Society*, 22(7), 900–915.

Hoffmann, A.L. (2020). Terms of Inclusion: Data, Discourse, Violence. [Preprint]. *New Media & Society*.

<https://static1.squarespace.com/static/5b8ab61f697a983fd6bo4c38/t/5f454d8b1ebedf389243f262/1598377356125/Hoffmann-TermsOfInclusionPreprint.pdf>

Holstein, K., Vaughan, J. W., Daume, H., Dudik, M., & Wallach, H. (2019). Improving fairness in machine learning systems: What do industry practitioners need? *CHI 2019*. <https://arxiv.org/pdf/1812.05239.pdf>

Honneth, A. (1995). *The struggle for recognition: The moral grammar of social conflicts*. Polity Press.

Hutchinson, B., & Mitchell, M. (2018). *50 years of test (un)fairness: Lessons for machine learning*. FAT* '19: Conference on fairness, accountability and transparency, Atlanta, GA. <https://dl.acm.org/citation.cfm?id=3287600>

Ibarra, H., Carter, N. M., & Silva, C. (2010, September). Why men still get more promotions than women. *Harvard Business Review*. <https://hbr.org/2010/09/why-men-still-get-more-promotions-than-women>

Irani, L. (2015, January 15). Justice for “data janitors.” *Public Books*. <http://www.publicbooks.org/justice-for-data-janitors/>

JafariNaimi, N., Nathan, L., & Hargraves, I. (2015). Values as hypotheses: Design, inquiry, and the service of values. *Design Issues*, 31(4), 91–104.

Katell, M., Young, M., Dailey, D., Herman, B., Guetler, V., Tam, A., Bintz, C., Raz, D., & Krafft, P. M. (2020, January 27–30). *Toward situated interventions for algorithmic equity: Lessons from the field* [Paper presentation]. FAT* '19: Conference on fairness, accountability, and transparency, Barcelona, Spain.

Keyes, O. (2018). The misgendering machines: The trans/HCI Implications of automatic gender recognition. In *Proceedings of the ACM on human-computer interaction – CSCW*. <https://dl.acm.org/citation.cfm?id=3274357>

Li, X. (2018). *Divination engines: A media history of text prediction* [Unpublished doctoral dissertation, New York University].

MIT. (1983). *Barriers to equality in academia: Women in computer science at M.I.T.* Prepared by female graduate students and research staff at the Laboratory for Computer Science and the Artificial Intelligence Laboratory at M.I.T.

Light, J. (1999). When Computers Were Women. *Technology and Culture*, 40(3): 455-483.

Mohamed, S., Png, M. T., & Isaac, W. (2020). Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. *Philosophy and Technology*. <https://doi.org/10.1007/s13347-020-00405-8>

Noble, S. U. (2012). Missed connections: What search engines say about women. *Bitch Magazine*, 12(54), 36-41.

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. NYU Press.

Oldenziel, R. (1997). Boys and their toys: The Fisher Body Craftsman's Guild, 1930-1968, and the making of a male technical domain. *Technology and Culture*, 38(1), 60-96.

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown Books.

Onuoha, M. (2017). *Notes on algorithmic violence*. <https://github.com/MimiOnuoha/On-Algorithmic-Violence>

Passi, S. & Barocas, S. (2019, Jan. 29-31). Problem Formulation and Fairness. [Paper Presentation]. Conference on fairness, accountability, and transparency (FAT* '19), Atlanta, GA. <https://arxiv.org/abs/1901.02547>

Phillips, A. & Taylor, B. (1980). Sex and skill: Notes towards a feminist economics. *Feminist Review*, 6, 79-88.

Rankin, J. (2018). *A people's history of computing in the United States*. Cambridge: Harvard University Press.

Raji, I.D. & Buolamwini, J. (2019, Jan. 7-8). *Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial AI products*. [Paper Presentation]. AIES'19, Honolulu, HI. <https://dl.acm.org/doi/pdf/10.1145/3306618.3314244>.

Richardson, R. (2020). Democracies Die in Black Boxes: How the History of US Civil Rights and Anti-Discrimination Laws Affects Algorithmic Bias of Today. Manuscript in Preparation.

Roberts, S. T. (2016). Digital refuse: Canadian garbage, commercial content moderation and the global circulation of social media's waste. *Wi: Journal of Mobile*

Media, 10(1), 1–18. <http://wi.mobilities.ca/digitalrefuse/>

Roberts, S. T. (2019). *Behind the screen: Content moderation in the shadows of social media*. Yale University Press.

Rose, J. (2019, August 13). Amazon says the face recognition tech it sells to cops can now detect "fear." *Motherboard*. https://www.vice.com/en_us/article/7x59z9/the-facial-recognition-system-amazon-sells-to-cops-can-now-detect-fear

Selbst, A. D., boyd, d., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2018). *Fairness and Abstraction in Sociotechnical Systems*. FAT* '19: Conference on fairness, accountability and transparency, Atlanta, GA. <https://www.ssrn.com/abstract=3265913>.

Spertus, E. (1991). *Why are there so few female computer scientists?* MIT Artificial Intelligence Laboratory Technical Reports.

Steele, C. (2018, January 4). The real reason voice assistants are female (and why it matters). *PC Magazine*. <https://www.pcmag.com/commentary/358057/the-real-reason-voice-assistants-are-female-and-why-it-matt?source=autosuggest>

Strok, D. (1992). Women in AI. *IEEE Expert*, 7(4), 7–22.

Sweeney, L. (2013). Discrimination in online ad delivery. *Communications of the ACM*, 56(5). <https://arxiv.org/abs/1301.6822>

Sweeney, M. (2018). The intersectional interface. In S. U. Noble & B. Tynes (Eds.), *The intersectional internet* (pp. 215–228). Duke University Press.

Suchman, L. (2003). *Located accountabilities in technology production*. Center for Science Studies, Lancaster University.

Taylor, A. (2018). The automation charade. *Logic Magazine*, 5. <https://logicmag.io/failure/the-automation-charade/>

Taylor, C. (1992). *Multiculturalism and the politics of recognition*. Princeton University Press.

UNESCO. (2019). *I'd blush if I could: Closing gender divides in digital skills through education*. <https://en.unesco.org/ld-blush-if-i-could>

Wacjman, J. (1991). *Feminism confronts technology*. Penn State University Press.

Wang, J. (2009, May 13). Racist camera! No, I did not blink...I'm just Asian! *JozJozJoz*. <http://www.jozjozjoz.com/2009/05/13/racist-camera-no-i-did-not-blink-im-just-asian/>

West, S. M., Whittaker, M., & Crawford, K. (2019). *Discriminating systems: Gender, race, and power in artificial intelligence*. AI Now Institute. <https://ainowinstitute.org/discriminatingystems.html>

Vincent, J. (2016, March 24). Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day. *The Verge*.

<https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist>

Vincent, J. (2017, August 22). Transgender YouTubers had their videos grabbed to train facial recognition software. *The Verge*.

<https://www.theverge.com/2017/8/22/16180080/transgender-youtubers-ai-facial-recognition-dataset>

Young, I. M. (1990). *Justice and the politics of difference*. Princeton University Press.

Young, I. M. (1997). Unruly categories: A critique of Nancy Fraser's dual systems theory. *New Left Review*, 222.

Zdenek, S. (2007). "Just roll your mouse over me": Designing virtual women for customer service on the web. *Technical Communication Quarterly*, 16(4), 397–430.

Author Bio

Sarah Myers West is a postdoctoral researcher at the AI Now Institute at New York University working on issues of gender, race, and power in AI. She received her doctoral degree from the Annenberg School for Communication and Journalism at the University of Southern California, where her dissertation examined the cultural history and politics of encryption technologies from the 1960s to the present day.