

Abstracted Away: Resisting Alienation and Ungrounded Abstraction in AI Research Communities

Vyoma Raman^{*1}, Isabel O. Gallegos^{*2}, Neha Srivathsa^{*3}

¹Department of Information Science, Cornell University, USA

²Department of Computer Science and Stanford Law School, Stanford University, USA

³Department of Computer Science, Stanford University, USA

vyoma@infosci.cornell.edu, iogalle@stanford.edu, nehasriv@stanford.edu

Abstract

Logics of abstraction in computational AI research often push important forms of knowledge and reflection aside: dominant standards of legitimacy become separated from lived experience of harm; the stated goals of work misalign with the practices that operationalize them; and career demands crowd out critical reflection. Even as prior efforts—both academic and community-oriented—have sought to recontextualize and challenge common practices, recurring exposure to sociotechnical harms and epistemic injustice has persisted. As three early-career critical AI researchers, we have experienced this as *alienation*: feeling like outsiders in our research communities. This process of alienation has involved having some aspects of our backgrounds overlooked and others tokenized. We argue that our alienation occurred through mechanisms that mirror abstraction by creating distance from relevant material realities. Beyond abstraction’s role in computational AI research as a foundational practice that structures complex computational tasks, we have encountered it as a social norm in computational research spaces, which we illustrate through an autoethnographic inquiry into our alienation. We narrate three vignettes that describe how we encountered and resisted alienation in our research communities. By analyzing themes across these accounts, we construct an interpretive framework of alienation that categorizes its preconditions, mechanisms, and harms. Finally, we identify affect abstraction, one of the mechanisms of alienation we describe, as a high-leverage mechanism that we can resist by staying attuned to our affective responses, and collective action as a way to reduce risk and isolation when engaging in resistance. To assist others with similar reflection, we present our framework as a hermeneutic resource. Engaging critically in self-reflection and meaning-making is a necessary step toward challenging exclusionary disciplinary norms and cultivating more inclusive forms of AI research.

1 Introduction

In many introductory computer science (CS) classes, students learn to write their first functions. They are taught to name a procedure, specify its inputs and outputs, and — crucially — to set aside what happens inside other functions it calls. For beginners, this is often presented as a practical convenience, but embedded in this simple practice is

a consequential idea. In CS, abstraction is the systematic practice of managing exposure to information about computational processes (Colburn and Shute 2007). This is useful and necessary for tasks like writing functions, designing programs, and learning to model patterns in data. More broadly, abstraction reflects values core to computing, such as using modular design, structuring experiments, and specifying constraints of potential interventions. In this work, we define abstraction as both **elimination**, the omission of unnecessary details, and **essentialization**, the emphasis of important details, to model a phenomenon *in service of a specified goal*. Both acts involve making some information more visible and other information less so, a framing highlighting that abstraction is inherently laden with value judgments regarding what kinds of details are important or unimportant for a given purpose. As these judgments are operationalized across systems and practices, they can become invisible and shielded from questioning (Bowker and Star 2000).

Our investigation of the mechanisms of abstraction is an attempt to understand our shared experiences of **alienation** in computational AI research spaces. We define alienation as the implicit and explicit dissonance that comes with feeling out of place.¹ It is an emotional response to a perceived absence of **intellectual safety** in academic contexts, which includes having one’s ideas and research valued and trusting others’ intentions. Our observations of sociotechnical harm first triggered this alienation, when we experienced grief and indignation at witnessing the violences caused by AI systems, especially toward minoritized communities.² We also noticed that, in some cases, those most vulnerable to harm from AI systems shared some of our personal characteristics, which seemed obvious to us but unapparent to our colleagues. When challenging the design decisions causing

¹We intentionally select this terminology to invoke alienation from one’s labor (Marx 1978). Although our definition differs, both forms of alienation occur when people become estranged from the systems and environments where they work.

²Drawing on multiple scholarly traditions (Dotson 2011; Farmer 2004; Galtung 1969, 1990; Nixon 2011; Scheper-Hughes and Bourgois 2004; Spivak 1988), we use the term “violence” to include physical, structural, cultural, epistemic, and other types of injustice and harm caused by sociotechnical systems, institutions, and asymmetries of power, while recognizing that these violences are not equivalent.

^{*}All authors contributed equally.
Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

these harms, we were subjected to various forms of **epistemic injustice**, which cast doubt on whether we belonged in computational AI research spaces. As put forth by Fricker (2007), epistemic injustice delegitimizes people in their role as knowers, occurring when social power and prejudice distort how knowledge is produced, shared, or understood.

These experiences led us to search beyond computational AI to other disciplines, including law, disability justice, education studies, and health policy.³ In these fields, particularly their critical branches, we encountered alternate traditions that sensitized us to how judgments about which details matter can have significant real-world consequences. Later, we attributed the alienation we had experienced to the misalignment we perceived between which details were simplified across settings and the communities those simplifications were meant to serve. We do not reject abstraction as a whole; rather, we highlight the tensions that arise when an impulse to engage in abstraction is pursued without grounding in purpose. We use **grounded abstraction** to refer to practices that are intentional about which details are eliminated and preserved, considering the context of the abstraction. In contrast, **ungrounded abstraction** refers to practices that strip away details critical to the phenomenon being modeled, rendering the abstraction ineffective for the stated goal or creating material risks and disproportionate disadvantage in the stated context. We provide these and other key definitions in Table 1, which we contextualize in Section 6.

Our holistic experiences working on computational AI research are critical to our inquiry: they have enabled us to recognize signals of alienation and name the mechanisms that drove it. At the time of writing, we are graduate students at private universities in the U.S., trained in CS and working on interdisciplinary, critical machine learning and AI research. To synthesize our observations, we employ an autoethnographic method. The vignettes we present, narrated in a collective voice, are evidentiary material that we analyze to identify how ungrounded abstraction has alienated us. In the style of Star and Bowker (2007), we blend the personal and analytic to formalize how ungrounded abstraction has driven our alienation in computational AI research spaces.

1.1 Roadmap and Contributions

In the following sections, we narrate our experiences navigating AI research cultures and analyze these to identify how alienation manifests in our experiences and how we have resisted it. In Section 2, we contextualize our autoethnographic method within a larger tradition of reflexive research practice. In Section 3, we recount our first vignette: an encounter with the ethically fraught yet often unquestioned practice of data cleaning in an undergraduate class. In Section 4, our second vignette, we depict experiences from the beginning of our graduate programs as we attempted to question assumptions and values underlying how computa-

³Our explorations beyond CS were made necessary by early challenges with identifying mentors to guide us on working on AI while staying aligned with our values. Interdisciplinary researchers face numerous challenges thriving within academia (Berkas et al. 2024; Mäkinen, Evans, and McFarland 2025; Zheng et al. 2025).

tional AI research should interact with the world. In Section 5, our third vignette, we highlight our experience engaging in an academic community where we found intellectual safety. While these vignettes touch on substantive issues including ethics education, discarding marginalized data points, human rights impacts of sociotechnical systems, and surveillance, we do not seek to intervene in each of them specifically. Rather, in Section 6, we juxtapose the thematic similarities and differences in our experiences across these vignettes, creating a framework that characterizes how our alienation has operated. Specifically, we argue that **our alienation occurred through mechanisms that mimic the logic of abstraction by creating distance from relevant material realities**. Finally, in Section 7, we derive concrete interventions and broader strategies of resistance from our vignettes, our framework, and prior literature.

We ground our work in scholarship on reflexive methodologies (§ 2) and prior efforts to synthesize structural issues in AI research (§§ 6.1-6.2). We also review prior, formally documented efforts to resist alienation within computational AI research spaces (§ 7), while acknowledging that much of this work lies in lived practice. This landscape shapes the contributions we claim:

1. We describe our experiences of alienation and of finding solidarity in three vignettes (§§ 3-5). These are evidentiary data and also acts of testimony.
2. We present an interpretive framework of our alienation, connecting the preconditions, mechanisms, and outcomes of ungrounded abstraction (§ 6). By highlighting structure that may be less visible in isolated accounts, we offer an interpretation of our alienation and provide a resource to help others engage in their own sense-making. We call back to prior literature in Sections 6.1-6.2 that relate to the specific phenomena that triggered our alienation.
3. We identify collective action and affective attunement as strategies to resist alienation and ungrounded abstraction. We frame this project itself as an intervention through the design of our methodology and output (§ 7).

2 Approach

Building on the traditions of feminist standpoint theory and critique of science, we take our experiences as an analytic resource. Feminist standpoint theory recognizes that marginalized social positions have unique perspectives on power and knowledge (Anderson 1995; Collins 1989; Harding 1992a,b). Situated knowledges make it possible to examine power relations (Haraway 1988; Lugones and Spelman 1983). The resulting insights can be applied to critique scientific knowledge (Adam 1993, 2000; Code 1991; Daston and Galison 2007). Recognizing injustice as contestable makes it possible to forge alternative futures (Freire 1970).

Computational AI research cultures have long lacked intentional reflexive practice, but there have been a variety of attempts to close this gap. In 1997, Agre (1997) called for a **critical technical practice** that interrogates the premises and practices of AI technologies. Agre and others have autoethnographically recounted and analyzed their experiences

Term	Definition
Abstraction	The superset of <i>elimination</i> , the omission of detail, and <i>essentialization</i> , the emphasis of detail.
Alienation	The emotive experience of feeling out of place due to a lack of intellectual safety.
Intellectual Safety	The set of epistemic and relational conditions required to participate meaningfully in scholarship.
Ungroundedness	The decoupling from real-world context and lived experience without justification, rendering something useless for the stated goal or generating material risks and disproportionate disadvantage in the stated context.
Metaeugenics	A set of logics that impact how nonconforming bodies are understood, treated, and regulated into compliance.
Legitimacy	The quality of those whom the field recognizes as a “real” AI researcher and the kinds of work it values.
Busyness	The demand of constant productivity and speed.
Testimony abstraction	The stripping away of one’s legitimacy to contribute to knowledge-making.
Purpose abstraction	The misalignment of the goal of pursuing a research project with how it is operationalized or evaluated.
Position abstraction	The flattening of social relations, power asymmetries, and differences in lived experience in data and AI systems.
Affect abstraction	The emotional distancing of a researcher from their work to better tolerate the harms they witness.
Epistemic injustice	The undermining of someone in their capacity as a knower.
Sociotechnical harm	Adverse effects toward affected communities in the social contexts in which AI systems are embedded.
Affective attunement	The practice of intentionally noticing and responding to one’s emotions as signals of real harm.

Table 1: Concepts that have structured our analysis of alienation in computational AI research spaces.

to provide insight on the barriers and pathways to recontextualize AI and CS research (Agre 1997; Hofmann et al. 2020; Khan et al. 2025; Russo et al. 2024; Suchman, Gururaja, and Widder 2025; Ymous et al. 2020). Other efforts have included applying critical theories to address existing practices (Hampton 2021; Hanna et al. 2020; Keyes, Peil, and Barlas 2019; Mohamed, Png, and Isaac 2020; Shew 2023), theorizing how technical practitioners develop critical perspectives (Malik and Malik 2022), and encouraging reflection in research outputs (e.g., Beygelzimer et al. 2021; NAACL 2022 Organizing Committee 2021; Olteanu et al. 2023). However, some analyses of these reflective practices suggest that authors insufficiently engage with potential impacts of their work, responsibility for harm, and how their positionality influences research outputs (Liu et al. 2022; Schroeder, Pareek, and Barocas 2025). Human-computer interaction research has also increasingly prioritized reflexivity in methods design and analysis (e.g., Cambo and Gergle 2022; Liang, Munson, and Kientz 2021).

Our work builds on these efforts by reflexively examining our experiences in computational AI research communities, and producing an artifact that assists others in doing the same. Toward the former, we adopt an autoethnographic method because it explicitly centers groundedness and the specificity of individual experience. By centering situated lived experience, autoethnography operationalizes feminist standpoint epistemology and enshrines affect, shifting away from a “view from nowhere” whose emphasis on neutrality and detachment makes abstraction seem invisible and therefore inevitable (Haraway 1988; Nagel 1986). As researchers immersed in computational AI research spaces, we can recognize the alienating effects of ungrounded abstraction on ourselves and analyze how these impacts occur.

2.1 Method

In autoethnography, researchers analyze their situated experience for insight into the conditions impacting them (Ellis, Adams, and Bochner 2011; Ellis and Bochner 2000; Ngunjiri, Hernandez, and Chang 2010; Stahlke Wall 2016),

including in relation to technology (Bala et al. 2023; Kaltenhauser, Stefanidi, and Schöning 2024). While autoethnography is often understood to emphasize emotive experiences (Ellis and Bochner 2000), some interpretations prioritize systematic analysis of researchers’ personal data (Anderson 2006). In our attempt, we identify similarities among our collective lived experiences and synthesize them to assist others with the same. We employ collaborative autoethnography, which performs self-interrogation within a group (Bundy et al. 2023; Chang, Ngunjiri, and Hernandez 2016). Some forms of collaborative autoethnography focus on individual reflective writing (Ngunjiri, Hernandez, and Chang 2010) and others emphasize group discussions (Noel, Minematsu, and Bosca 2023); we used both approaches.

Our data collection process unfolded over six months, during which we recorded 16 hours of discussion focused on our experiences in computational AI research spaces and our interpretations of them. We summarized these conversations in detailed written notes and engaged in an iterative, inductive thematic analysis. This follows collaborative autoethnographic approaches (Kafar and Ellis 2014; Nel 2018; Ngunjiri, Hernandez, and Chang 2010; Rutter et al. 2023). Each author independently conducted an open coding pass of the written notes to identify themes and produced short analytical memos synthesizing them in different ways. Following this, we had a series of discussions to reach consensus on the higher-level thematic groupings in the memos. From this analysis, we assembled three vignettes of our encounters with alienation that maximized coverage of the most salient themes. By articulating experiences that previously felt isolating and recognizing their shared structure, we lessened the alienation we had each carried. To reflect this experience, we describe our findings through vignettes narrated in a collective voice, representing multiple authors’ accounts while remaining grounded in specific moments.

3 Vignette: We Clean Ourselves from Data

We had a formative encounter with the dynamics of alienation and abstraction in an introductory data science course.

As part of an assignment, we walked through data cleaning procedures to use medical records to model insurance risk scores. One step, presented as routine, required us to remove data points flagged as anomalous — people whose data features deviated too far from the statistical norm. These cases were labeled noise. Excluding them, we were told, would clarify the “real” signal that the model was supposed to learn about the overall population, and the assignment stated that including them would make the model worse. In practice, this meant discarding people whose lived realities and medical histories did not align with the majority, reinforcing whose lives were worthy of prediction and whose were not.

It did not escape our notice that we were part of the groups being removed. We occupied the statistical tails that the assignment framed as expendable: people whose biomarkers fell outside general standards and whose identities were “Other.” The assignment forced us to enact an eliminative logic to filter out people like us. Rather than neutral preprocessing, we were being taught to routinely operationalize a logic of data “cleaning” that disciplined deviance.

A discussion question prompted us to reflect on the people represented in the data being dropped, recognizing that their removal could lead to worse model performance and downstream adverse effects for the groups they represented. But by proceeding with the “cleaned” dataset anyway, the assignment indicated that this should not alter the workflow. The underlying lesson: we were an unfortunate but inevitable tradeoff, since the model’s performance for people like us ultimately did not matter. The brief interlude to discuss an ethical issue was just that — an interlude — sufficient to satisfy the instructional goal of teaching ethics while rendering followup changes to the approach unnecessary.

The stated purpose of this assignment was to teach trainee data scientists the best practices for effective predictive modeling, yet it offered no guidance on making principled tradeoffs between representing a diversity of people and isolating a “clean” pattern. By avoiding discussion of how to balance these competing aims, the assignment reduced a complex judgment to a simple rule in an act of ethics-washing. This left us with little intuition on how to apply the lesson to a realistic scenario. Our primary learning was that grounded reasoning could be acknowledged and then dismissed, and that contextual judgment was unnecessary.

Unable to contest a pre-programmed autograder, we had no choice but to rationalize this approach and suppress our discomfort. Data cleaning had been introduced to us as a necessary process to help models perform well on a desired population, and as students, we assumed that our unease reflected our inexperience rather than a substantive issue with the methodology. We trusted our professors to have considered perspectives like ours when designing the curriculum.

Later assignments in the course reinforced the learning from this one: the final project involved a leaderboard where credit was distributed according to standard performance metrics on a hidden dataset and where any approach was allowable. To succeed, we had to disregard potential downstream effects our methods might have on the individuals represented in the data. Under pressure to complete heavy projects in a tight timeline, it was surprisingly easy to file

away our concern to revisit later. Thus, by decoupling substantive ethical reflection from the criteria judging our work, our professors abstracted away who the model represented and how. This left a decontextualized notion of model performance, not human impact, as the definition of success.

When the busyness of classes slowed and we regained our capacity, the marginalization embedded in the practices we had learned became clear. We had been taught to accept the exclusion of people like us as a “reasonable” sacrifice for the greater good, internalizing that our existence in the training data was incompatible with a useful model. This interpretation positioned our exclusion as computationally justified, a logic that extended far beyond us. No doubt similar justifications had been used in real-world contexts. The groups most likely to be removed from such datasets during data cleaning steps (such as disabled people, queer people, and racial and ethnic minorities) are also those most under-served by and vulnerable to many social systems. That is, the individuals abstracted away from the relevant AI systems would face greater material risk from any errors than the ones left in. By ignoring the impacts of flawed algorithms on people removed from the training data and adopting an approach that increased the likelihood of disproportionate predictive errors affecting those same groups, our introductory assignment implied that those most likely to face harms from a system need not be meaningfully represented within it.

4 Vignette: Conflicting Visions for Computer Vision

In the first month of our graduate degrees, two encounters shaped how we experienced our research environments. The first was a significant escalation of colonial and imperialist violence, which we encountered through social media, journalistic reporting, and conversations with those who were personally affected and organizing to protest it. Media coverage increasingly brought to light how AI technologies, including computer vision (CV) tools, were being used for military purposes. The second encounter was a paper that traced the connection between CV research and downstream surveillance applications (Kalluri et al. 2025). This work illuminated how CV research conducted in academic and other settings, including application-agnostic technologies or those built for beneficial and benign applications, can simultaneously be used for oppression. It sickened us that efforts to build AI for social good could be corrupted like this — and that such dual uses have characterized the field since its beginnings (Leslie 1993; Selinger and Durant 2022; Waelen 2024). Even as reports on such destructive applications became widespread, we observed that they were still overlooked in the CV research discussions around us, which emphasized potential benefits over associated risks.

These threads began to overlap. Our peers drew attention to how surveillance technologies were already all around us, especially in areas of campus most frequented by minoritized students. We observed the embodied effects of these surveillance technologies: we changed our movement patterns, were hyper-aware of cameras and potential surveillance devices when traversing campus, and carried tension

that we didn't before. Activists pointed to the use of CV in the ongoing imperialist escalation, heightening our awareness of how research enabled the devastating violence reported in the news. These observations raised unavoidable questions about the research that we were conducting and the ways that the research careers we had at one point envisioned for ourselves could enable oppressive regimes.

At the time, we were each working on advancing AI capabilities in different areas, hoping to leverage technology to improve lives. Facing the reality that many AI systems are implicated in violence, we questioned whether the research agendas that we were contributing toward would enable oppression (Kalluri et al. 2025; Stark 2019; Suchman 2015). We worried that working on research to improve AI would require complicity. These questions weighed on us heavily. We found ourselves bringing up these topics in conversation, hoping to understand our peers and mentors' mental models of impact and how they reconciled ethical dilemmas.

Among fellow graduate students, we saw curiosity about the social impacts of research, tempered by an emotional distance. We felt that these concerns were more central to our own day-to-day lives in academia than our peers'. Like others, we experienced a constant stream of research, teaching, and administrative tasks, and faced the speed of publishing cycles in AI. But despite pressures to gain credibility as early-career researchers, working on specific topics with specific people at the cost of decentering ethics from our research agendas was not a tradeoff we were willing to make.

When discussing our dissonances with more established researchers, we received guidance illustrating the constraints of dominant research agendas. We were encouraged to build "ethical" versions of AI technologies. We saw this as reflecting an underlying belief that developing such tools, as currently envisioned, was inevitable, and that ethical responsibility mainly consisted of incremental improvements. While aligning with these agendas would undoubtedly make certain career paths, such as CS academia, more attainable, we were concerned that this would require us to decenter the values that we hoped to advance through our research, which prompted our interest in graduate school in the first place. Although we acknowledged the care that this advice stemmed from, we were in search of guidance on how to resist the pressure to conform to prevailing research agendas without being perceived as naive and dismissed as scholars.

In considering the mission of developing more ethical versions of AI technologies, some researchers cautioned us against being positioned as "merely" the "ethics person" in our respective areas, warning of tokenization and minimization of our computational skills. These comments led us to question whether critical work in these spaces truly had the power to transform dominant computational agendas, or whether it would only be granted legitimacy when affirming existing research agendas. These reflections felt sobering.

We had hoped that these conversations would support our collective growth alongside our colleagues and help us refine our research goals. While we felt cared for, we left unsure how to proceed. The questions we were asking felt impermissible in many research spaces around us, and the responses we received often reflected assumptions and val-

ues about AI development that we disagreed with. These premises felt difficult to escape. We were left longing for a community to explore such questions with.

5 Vignette: Intellectual Safety in Community

Upon feeling alienated from computational AI research communities, we began searching for alternate spaces where our concerns would be constructively centered as respectable research inquiry. We sought to critique computational advancements and envision liberatory computational futures in intellectually safe environments. We experienced some successes in finding community: the authors met each other and identified mentors who had had similar experiences. Some mentors introduced us to others, but eventually, despite our desire for more intellectual companionship, our networks of critical scholars in CS stopped growing.

During this period, we found a community of critical technology scholars based out of Stanford's Graduate School of Education. There, we found peers grappling with similar topics from diverse disciplinary vantage points including, but not limited to, CS. When a paper describing the eugenic ideologies underlying modern AI (Geburu and Torres 2024) was released, one coauthor volunteered to host a reading group session within this community to discuss this paper. RSVPs, which included the other two coauthors, spanned scholars in CS, education, science and technology studies (STS), communication, and sociology. The ensuing event was lively: discussion topics ranged from dissecting the paper's argument to analyzing its discursive intervention. Due to the diverse disciplinary backgrounds in the group, each attendee brought a different perspective and area of expertise. Each person had to translate the methods, concepts, assumptions, and ideologies of their discipline for others, intentionally abstracting away some context to make their point legible.

After the event, the three authors walked to our workspaces together. As we did so, we mused about why this reading group had felt so positive, in contrast to the alienation that we persistently felt in computational AI research spaces. We each experienced an embodied sense of ease in the space, so distinct from the tension we often carried in CS spaces. We identified three notable differences. First, we noticed the pace of conversation was slowed; people were respectful of others' speaking time, left space to reflect, and engaged with what others had shared. Second, due to the interdisciplinarity of the group, participants did not take assumptions for granted, instead spending time to establish a common knowledge base. Third, it was not only acceptable but welcome to critique the values and assumptions of the work. Altogether, these norms created a markedly different intellectual experience and affirmed our desires for more.

Our conversation during this walk was the first time we had discussed our alienation as a group rather than in pairs. Identifying the parallels between our experiences helped us validate our feelings of alienation. Further, we became aware that frustrations, which previously felt like isolated experiences or coincidences, might be caused by structural, cultural, and institutional factors. We began to consider that,

beyond merely coping with our alienation, we could examine and even actively resist the factors impacting it.

At the time, the three of us were at different transition points: some nearing the end of PhD rotations and deciding next steps, and some deciding which next degree programs would serve our intellectual needs. We sought to replicate the microcosm of our discussions on that day at a greater scale. We were motivated by the intuition that being part of a larger community that does careful and critical AI work would help us become the scholars we wanted to be and do the kind of research that would push AI in the directions we wanted. This conversation between the three of us, where we affirmed each of our respective goals, identified the structural factors driving our experiences, and considered how logics of eugenics may permeate AI systems and communities, was, in retrospect, an early inception of the work that forms this paper.

6 Analyzing Our Alienation

We examine the factors that created or averted our alienation in each vignette, finding them to follow similar patterns that we characterize in this section.

6.1 Alienation in Context

We experienced alienation as an emotional reaction to recurring dynamics experienced across coursework, meetings with collaborators, and other contexts. The contrast between our experiences of alienation and the lack thereof is depicted in the previous sections and expanded upon in the Appendix (<https://arxiv.org/pdf/2608.08408>). These reveal underlying tensions that perpetuate alienation that have been extensively theorized by various literatures.

Interdisciplinarity Prior literature has explored the challenges of legibility in interdisciplinary research. Interdisciplinarity is celebrated for its “essential tension” of bringing together different fields for disruptive innovation (Kuhn 1977). It requires translating between epistemologies and methodologies, intellectual values and cultures, and organizational and institutional traditions (Bauer 1990; Fahimi et al. 2024). Interdisciplinarity is harder still in the context of computational AI research, which has particular tensions with humanistic and socially-oriented fields (Fahimi et al. 2024; Klein et al. 2025). These tensions include demands for formalization over open-ended interrogation; fast progress on well-defined problems over slower inquiry; and generalizability over contextual grounding (Klein et al. 2025; Klumbyte, Draude, and Taylor 2022). Beyond creating mutual friction between fields, interdisciplinary work can seek to avoid challenges to dominant perspectives, leading to systematic marginalization of alternative approaches (Klein et al. 2025).

Institutional Logics in AI Ethics The struggles in our vignettes emerged when tensions that felt unavoidable to us clashed with institutional logics that deemed our concerns as unimportant. Prior literature has described how dominant AI ethics work is institutionally situated, influenced by a combination of corporate, academic, and political power (Bietti

2020; Green 2021; Metcalf, Moss et al. 2019; Young, Katell, and Krafft 2022). This impacts what is recognized as legitimate AI ethics work and who is able to conduct this work. Under institutional logics, AI ethics can become oriented toward legitimizing AI advancement rather than meaningfully influencing or constraining it (Green 2021).

Minoritized communities are systematically underrepresented in AI ethics, in the consideration of harms (Birhane et al. 2022b) and in their participation in shaping the field. AI researchers may define ethics in ways that strip context away (Selbst et al. 2019), leading to obscured power differentials that impact minoritized communities. AI ethics work that is based in lived experience, often conducted by people from minoritized groups, is delegitimized, while other AI ethics work uses quantification as a strategy for legitimacy (Widder 2024). This privileging of computational ways of knowing reproduces epistemic exclusion even at the level of CS education (Raji, Scheuerman, and Amironesei 2021).

Legitimizing Situated Knowledges Tensions around which situated knowledges are legitimized or dismissed persist, even as many venues encourage the integration of positionality and reflexivity into research outputs (AAAI AIES 2026; ACM FAccT 2026). Formal positionality statements have been found to insufficiently connect to research outputs, and they are still interpreted and implemented inconsistently (Schroeder, Pareek, and Barocas 2025; Singh et al. 2025). Empirical findings further complicate this: while these venues have diversified the research topics they publish, they continue to reproduce structural inequities in whose knowledge is represented (Acuna and Liang 2021). This suggests that researchers differ in their ability and desire to reflexively change their research practices in the ways that such measures seek to promote. As such, much work remains to legitimize diverse knowledges in AI research.

6.2 From Alienation to Abstraction

The vignettes illustrate our experiences of alienation as not just a lack of care or a mismatch between our values and those of our research communities, but as a product of pressures we faced to conform to a research agenda that prioritized contributions to AI development over those to society, elevated potential benefits of technology over harms, and emphasized optimization on generalizable problems while leaving underlying normative commitments unstated or mis-specified (Birhane et al. 2022a; Laufer, Gilbert, and Nissenbaum 2023). While we initially attempted to oblige, the stress of pretending that harm wasn’t occurring took a toll on us that did not alleviate until we began to address our complicity. We were further expected to enact modes of research that assumed a view of technological progress centered on values like generalization, efficiency, benchmark performance, and novelty on large datasets (Birhane et al. 2022a), pursuing our work only within them, when what we wanted was to interrogate the harms inherent to this status quo and imagine alternate technological futures. Sharing these experiences with each other, we began to realize their structural components and carve out a space of intellectual safety without these constraints.

This pressure toward conformity resembles a regime that Williams (2019) characterizes as **metaeugenic**: containing “cultural norms, ideals, values, and demands that warp and twist deviant bodies into conformity.” Encapsulating the logics of sanitization, normativity, and compliance, metaeugenics is a thread that summarizes our experiences of alienation and has been established in the contexts of computational cultures and systematic oppression (Chan 2025; Gebu and Torres 2024; Williams 2025). We apply metaeugenics to study how deviance in our computational AI research communities is disciplined into conformity.

Conformity requires stripping away one’s “aberrant” characteristics, an act reminiscent of abstraction’s process of omitting certain details. To understand how abstraction sustains a metaeugenic regime in computational AI research communities, we consider its normative nature. Much scholarship has questioned which and whose values are embedded in computational systems and how abstraction can obscure power asymmetries and harms (Laufer, Gilbert, and Nissenbaum 2023; Selbst et al. 2019; Wang 2025), reminiscent of earlier accounts of idealized formalism displacing embodied experience (Husserl 1970). In addition to influencing computing technologies, abstraction shapes CS by enabling researchers and developers to remain detached from the contexts where their outputs will be used, starting as early as introductory CS education (Birhane et al. 2022a; Malazita and Resetar 2019; Peterson, Ferreira, and Vardi 2023). More broadly, philosophers of science have emphasized that values shape both scientific research and its societal consequences (Douglas 2000; Longino 2020). Together, this work suggests that abstraction mediates values within computational research communities and in sociotechnical artifacts that impact the world. Even so, abstraction’s potential for harm or benefit depends on how much it is grounded in context and goals. Our vignettes make these impacts visible in computational AI research by showing how it, in our experiences, shapes not only research outcomes but also participation, legitimacy, and belonging.

6.3 Interpretive Framework of Alienation

We propose a framework (Figure 1) that links our experiences of alienation to patterns of abstraction, the conditions that sustain them, and the harms they produce. This framework reflects thematic similarities across our three vignettes that emerge from our coding of them.

Our framework contains three types of components: (i) **metaeugenic preconditions** that shape research culture; (ii) **mechanisms of ungrounded abstraction** that operate within it; and (iii) **harmful outcomes** resulting from these mechanisms that affect researchers and society. We connect each component to our experiences with an alienating example from one of the first two vignettes and a counterexample from the third, listing the corresponding vignette with each example. We also briefly characterize the preconditions and mechanisms using prior literature.

Metaeugenic Preconditions Drawing on prior work (§ 6.2), we identify metaeugenic preconditions as critical to shaping our experiences of computational AI research. We

frequently saw two particular norms enabling the abstracting mechanisms in our experiences:

- **Legitimacy** governs who the field recognizes as a “real” AI researcher and what kinds of work it values. Research that emphasizes methodological novelty, disciplinary purity, and alignment with dominant research agendas is seen as legitimate, while work studying lived experience, interdisciplinary critique, or adverse consequences within a context is often secondary or out of scope.

Prior literature: Different fields have distinct epistemologies and approaches that implicitly exclude certain ways of knowing (§ 6.1, Bauer 1990; Fahimi et al. 2024), and legitimacy can systematically foreclose alternate approaches (Klein et al. 2025; Klumbyte, Draude, and Taylor 2022). The institutional logics of AI ethics (§ 6.1) further legitimize harmful AI work rather than constraining it (Green 2021). We draw from these ideas that center power, where structural asymmetries determine whose research agendas are recognized as belonging to the field.

Example (§ 4): Mentors warned us about the organizational distinction some labs make between “technical” researchers who advance AI capabilities and “ethics” researchers who study or mitigate their oppressive uses.

Counterexample (§ 5): Event participants translated between different disciplines, enabling broader engagement in critical discussions of AI.

- **Busyness** demands constant productivity and speed. Grounded work, especially work that requires engagement with affected communities or ethical uncertainty, is often positioned as inefficient or unnecessary.

Prior literature: Institutional logics encode capitalist and corporate structures that promote metricized, productivity-oriented ways of working (Bietti 2020; Green 2021; Metcalf, Moss et al. 2019; Widder 2024) over slower, critical engagement (Klumbyte, Draude, and Taylor 2022). Our notion of busyness connects this to the pace demanded by these pressures.

Example (§ 3): The fast pace and heavy workload of our CS courses prevented us from engaging more deeply with the implications of the techniques we learned.

Counterexample (§ 5): Discussion proceeded at a pace that allowed participants to fully express their ideas and reflect before engaging, without fear of being left behind.

Mechanisms of Ungrounded Abstraction We identify four types of abstraction that reinforce the metaeugenic regime: testimony abstraction, purpose abstraction, position abstraction, and affect abstraction.

- **Testimony abstraction** concerns who is heard. Abstraction strips away one’s legitimacy to contribute to knowledge-making in computational AI research spaces. Researchers’ credibility is selectively amplified or diminished by framing their expertise or experience in ways that reify disciplinary boundaries. Critique is dismissed either because a researcher is deemed insufficiently “technical” or because their concerns are framed as naive, emotional, or inevitable. In these cases, critical

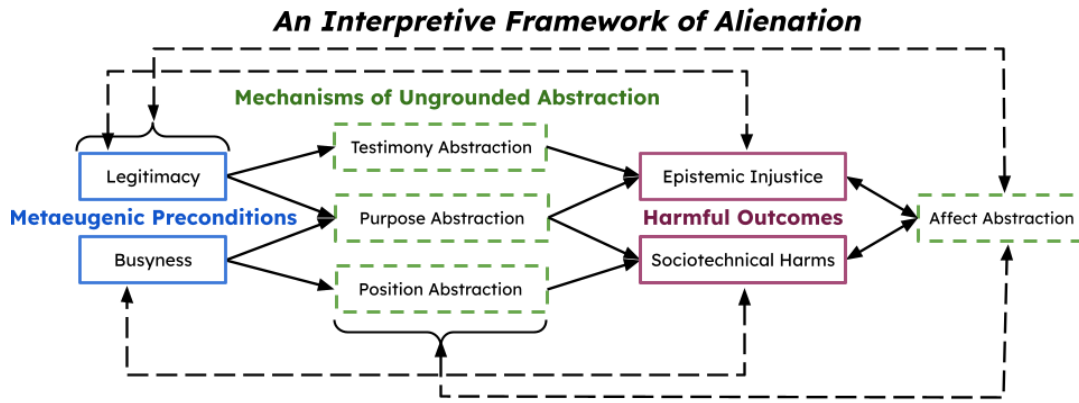


Figure 1: Interpretive framework of alienation, as present in our own experiences. We identify two metaeugenic preconditions that enable four mechanisms of ungrounded abstraction, which generate two categories of harmful outcomes. Arrows imply proposed causal relationships. Directed arrows indicate that the source node impacts the destination node, whereas bidirectional arrows indicate mutual influence between the two nodes. We highlight a pathway from preconditions to outcomes with solid arrows; dashed arrows represent how the components perpetuate each other.

scholars of computational AI research experience testimonial injustice (Fricker 2007).

Prior literature: Prior studies show that the work of researchers from minoritized communities is disproportionately delegitimized (Birhane et al. 2022b; Raji, Scheuerman, and Amironesei 2021; Widder 2024). Fricker (2007)’s account suggests that this is a structural reduction of those researchers’ credibility.

Example (§ 4): *We worried about being perceived as naive when we questioned the necessity of building certain AI technologies, based on their violent uses.*

Counterexample (§ 5): *Every researcher added value to the discussion due to their distinct background.*

- **Purpose abstraction** concerns why an action is taken. Abstraction occurs when the reason for pursuing a project becomes misaligned with how it is operationalized or evaluated. Methods may drift away from the problem they claim to address, and may deviate in ways that actively oppose the original intent. Career incentives, prestige, or novelty may impede substantive engagement with the original goal.

Prior literature: The purpose of research has been found to diverge from its methods when translation across epistemologies and institutional cultures breaks down during interdisciplinary collaborations (Bauer 1990; Fahimi et al. 2024; Klumbyte, Draude, and Taylor 2022). AI ethics frameworks may be applied superficially to serve the advancement of the research they are meant to scrutinize (Green 2021; Widder 2024).

Example (§ 3): *In our coursework, the goal of training ethical computer scientists was diluted when the assignment chose to merely name ethical issues instead of modifying approaches to address them.*

Counterexample (§ 5): *The group engaged with the underlying assumptions and values of the paper as well as*

its other components; this intentionally holistic critical discussion made space for wide-ranging examination.

- **Position abstraction** concerns who is affected. Abstraction removes social relations, power asymmetries, and differences in lived experience from data and sociotechnical systems. It may erase imbalanced power or assume uniformity and interchangeability between developers, users, and subjects of AI systems. Together, these obscure how marginalized groups may be disproportionately impacted in ways that exacerbate inequities.

Prior literature: Previous interventions to encourage reflexivity and analysis of situated knowledge (§ 6.1) include positionality statements specified by publishing venues (ACM FAccT 2026; AAAI AIES 2026). While these recognize that researcher positionality shapes research, it is difficult for an intervention that frames positionality as a disclosure problem to encourage active reflection on how one’s position and power meaningfully shape the research process (Liu et al. 2022; Schroeder, Pareek, and Barocas 2025; Singh et al. 2025). Institutional logics compound this by limiting access to who can meaningfully shape the field (Bietti 2020; Birhane et al. 2022b; Green 2021; Metcalf, Moss et al. 2019) and by removing the contextual specificity that makes disproportionate impacts visible (Selbst et al. 2019).

Example (§ 3): *We filtered from the training data people whose positionalities both made them candidates for removal and would uniquely make them vulnerable to harms from the AI system.*

Counterexample (§ 5): *The discussion’s chosen emphasis on eugenics foregrounded considerations of deviance and who is considered valuable.*

- **Affect abstraction** concerns how researchers feel about the effects of their works. It occurs when researchers emotionally distance themselves from their work to better tolerate the harms they witness without disrupting the

prevailing metaeugenic norms. Each of the other types of abstraction enables this distancing by reframing adverse consequences as a computational necessity or acceptable trade-off. By enforcing emotional discipline, affect abstraction renders metaeugenic norms as natural and the harms produced by other types of abstraction as inevitable. In contrast, **affective attunement** involves remaining responsive to emotions as sources of epistemic and ethical insight; by making harm emotionally visible, attunement destabilizes the metaeugenic regime in ways that are difficult to dismiss.

Prior literature: Affect theory broadly explores embodied responses to one's experiences, including how social power operates within environments (Clough 2007). Reflexive practice in research taps into these and other experiences as sources of knowledge (Hampton 2021; Hanna et al. 2020; Keyes, Peil, and Barlas 2019; Mohamed, Png, and Isaac 2020; Shew 2023). Our conception of affect abstraction emphasizes the role of affect in identifying mechanisms of injustice (Lorde 1981) and the harm that occurs in the absence of emotional attunement.

Example (§ 4): When we learned about CV applications that threatened human rights, we were encouraged to suppress our discomfort with the potentially harmful uses of the technologies we worked on.

Counterexample (§ 5): Collectively articulating experiences of feeling out of place helped us identify the recurring emotional patterns shaping our alienation.

Harmful Outcomes We connect these preconditions and mechanisms to two established classes of harmful outcomes that have perpetuated different violences on researchers and communities impacted by sociotechnical systems.

- **Epistemic injustice** causes researchers' knowledge and credibility to be unfairly dismissed (Dotson 2011; Fricker 2007). This arises from testimony abstraction that challenges the legitimacy of critics, and from purpose abstraction that evaluates researchers with metrics decoupled from their intellectual contributions.

Example (§ 4): We were wary of being seen as just "ethics people" in our labs and having our technical expertise discounted.

Counterexample (§ 5): Our community of critical technology scholars took seriously our contributions as both technical and ethical experts.

- **Sociotechnical harms** toward affected communities arise when AI systems produce adverse impacts in the social contexts in which they are used (Shelby et al. 2023). These occur through position abstraction that prevents AI practitioners from adequately understanding the social relations surrounding AI systems, and through purpose abstraction that permits methods that may conflict with the purported social goal of a system.

Example (§ 4): Militaries use computer vision weapons against populations subjected to colonization.

While this framework reflects how ungrounded abstraction perpetuated alienation in our own experiences, we emphasize that our goal is *not* to offer a generalized theory.

Rather, we offer this framework as a generative resource to assist others in making meaning of their own experiences. Given the shortcomings of some existing approaches to researcher reflexivity (Schroeder, Pareek, and Barocas 2025; Singh et al. 2025), leveraging this framework may help deepen reflection. We further describe the relationships between the nodes in our framework in the Appendix (<https://arxiv.org/pdf/2608.08408>).

7 Discussion

Ungrounded Abstraction Fuels Alienation. Our framework links moments of alienation from our research communities to the abstraction mechanisms through which they occurred. Each form of ungrounded abstraction eliminated or essentialized the contexts that grounded our work. As a result, our critiques lost salience. Much as notions of objectivity, neutrality, and the "view from nowhere" strip away the knower, each form of ungrounded abstraction removes one's accountability to situated experience. While our framework calls upon feminist standpoint epistemology, **our analysis of alienating experiences identifies a logic of abstraction within our computational AI research communities.** Ungrounded abstraction actively produces the conditions of alienation by removing testimony, purpose, position, and affect from everyday research practice. Thus, our alienation signaled when the presence of ungrounded abstraction pressured us to conform to a dominant research agenda.

Our critique is not unique to computational AI research cultures, but we argue that the field amplifies the dynamics of alienation in distinctive ways. Particularly, AI enables the rapid operationalization of abstraction into sociotechnical infrastructures (Selbst et al. 2019), and AI research cultures privilege quantitative formalizations over other forms of knowing, which can make contestable assumptions seem objective or inevitable (Adam 1993; Agre 1997).

Proposed Interventions By connecting ungrounded abstraction to metaeugenic preconditions, our framework of alienation indicates that there exist interventions that can act on both. We suggest avenues for this below⁴ and invite others to interact with our framework by using their personal experiences to identify new practices.

The accessibility of interventions may vary across career stages, reflecting differing power and risk. More established researchers are positioned to reject metaeugenic ideals of busyness and legitimacy. They can resist busyness by eschewing rapid publishing cycles that incentivize the sacrifice of depth and contextual grounding. In their labs, departments, and relevant research associations, they can encourage the questioning of disciplinary premises. To broaden standards of legitimacy within the field, they can platform a range of perspectives at events and use their power as reviewers or editors to affirm the value of topics or methods left out of dominant scholarly discourse. To create protective conditions for early-career scholars to engage in resis-

⁴Many factors affect the ability to change research practices. We draw on the framework and vignettes to propose interventions, so we are inherently limited by the scope of our method and do not claim our suggestions to be universal.

tance, more established scholars can mentor them on how to navigate institutional structures while encouraging them to question those structures and develop the skills needed to transform them. This could include creating transparency around how decisions influencing academic spaces are made and making visible the hidden curriculum of academia.

Early-career researchers, especially students, are primarily responsible for executing projects. This empowers them to avoid ungrounded abstraction in everyday research decisions. For instance, they can resist position abstraction by being intentional about how they operationalize groups and identities within datasets; testimony abstraction by centering scholarship from minoritized researchers; purpose abstraction by assessing whether design decisions align with the goal of their research projects; and affect abstraction through deliberate affective attunement, a strategy we detail below.

Even as we believe in the power of all people to transform research norms, the labor of resistance is often distributed unevenly. Students are particularly at risk for making research decisions that conflict with the direction of superiors, while remaining susceptible to distress from engaging in ungrounded abstraction. We encourage researchers at all levels to reflect on their positionalities and the forms of resistance accessible to them. To manage these risks, we identify collective action and affective attunement as key strategies.

Resisting Alienation Through Collective Action. We identify collective action as essential for resisting mechanisms of alienation, especially given our experiences in Section 5. Collective action can distribute the risks of deviating from norms. When early-career researchers exercise their power to challenge practices collectively, retaliation becomes more difficult. Approaches can include sharing experiences privately and publicly, coordinating pushback on harmful norms, sharing risk through joint statements, and creating visible spaces of collective critique, such as reading groups or workshops. Visible critical spaces can also serve as gathering points for organizing collective power. Communities in technology spaces are already engaged in important forms of collective action, including labor organizing efforts like the Tech Workers Coalition, identity-based communities like Queer in AI, data governance initiatives such as the Indigenous Data Sovereignty Network, academic collectives like the Stanford Critical AI Group, and geographically distributed events like Deep Learning Indaba.

In addition to critiquing harmful structures, transforming research practices requires envisioning and creating alternatives. This involves **futuring** and **prefiguring** desired research communities by enacting in the present the ethics and epistemic practices that guide the versions to be realized in the future (Boggs 1977). We participated in an iteration of this strategy (§ 5) by creating a community grounded in interdisciplinary critique. Further, as we experienced with each other, community support can enable shared interpretation of alienating experiences, which may otherwise remain obscured by affect abstraction. Community can protect researchers by distributing risk and creating the conditions for affective attunement that may be less accessible in isolation.

Resisting Ungrounded Abstraction Through Affective Attunement. Inspired by Lorde (1981) and Ahmed (2004), and building on the humanities’ “affective turn” that situates embodied reactions within systems of labor, technology, power, and control (Clough 2007), we call for **affective attunement**, the practice of intentionally cataloging one’s emotions as real signals of harm. Our experiences indicate that engaging in affective attunement can shield against affect abstraction and resulting alienation. Because the other components in our framework have bidirectional relationships with affect abstraction, it is an ideal intervention point.

Since affect abstraction creates distance between researchers and the harms of the norms they follow, affective attunement can reduce that distance. By centering researchers’ emotional discomfort with research practices, it makes space to question whether those practices are necessary, challenge their assumptions and values, and identify alternatives. We found that our experiences of alienation, particularly through epistemic injustice, were most acute when we failed to acknowledge our emotional responses. In these moments, we were subjecting ourselves to epistemic coercion; by succumbing to pressure to act in ways that contradicted our perceptions, we came to treat dominant interpretations of our circumstances as more legitimate than our own (Dandeleit 2021). By rendering harms visible, affective attunement expands which patterns of justification are ultimately considered reasonable. This makes it a powerful form of resisting ungrounded abstraction. Affective attunement complements other interventions that seek to amplify researchers’ lived experiences in CS research spaces through storytelling (Ogbonnaya-Ogburu et al. 2020; Ymous et al. 2020), centering the knowledge of those who experience harm (Birhane 2021; Keyes 2018; Scheuerman, Branham, and Hamidi 2018), and reflecting on one’s own positionality and accountability (Klumbyte, Draude, and Taylor 2022).

We recognize that centering affective attunement risks emphasizing individual correction over systemic accountability. This may shift responsibility onto marginalized people or those with less institutional power, burdening them with additional emotional and interpretive labor. We emphasize that affective attunement should operate collectively, such as in a research lab, where shared practices of collectively identifying and responding to emotional discomfort can establish solidarity and collective accountability.

Situating Our Work As Intervention. We see this work as an intervention to resist alienation that occurs from ungrounded abstraction. For us, the conversations we shared during analysis and consensus-building have helped build a community with intellectual safety. We adopted a methodology that explicitly resists epistemic injustice by claiming evidentiary value in our lived experiences. As the vignettes illustrate, academic papers by critical scholars also helped create the conditions for us to reflect on the academic tensions we experienced. Particularly critical for us was being able to name the structural factors driving these experiences. In the hope that we can do the same for others, we offer our interpretive framework as a hermeneutic resource for making meaning of one’s own experiences of alienation.

Positionality Statement

The meaning we construct in this work emerges as we occupy roles where we often feel like the only people with our particular lived experiences and perspectives within our research communities. We recognize that these positions create a mix of hyper-visibility and isolation that shapes how we perceive others and how others perceive us. Among the authors, all three identify as women. Two are Indian. One author is mixed-race and Hispanic. One is queer. One is disabled. We have all inhabited elite research institutions that pride themselves on being global leaders in AI innovation while remaining deeply steeped in and sustained by AI extractivism. Our work is therefore shaped by shared academic training rooted in institutional contradiction, where we learn from and gain power from the very systems we seek to resist. These experiences collectively shape our attentiveness to power and exclusion within the spaces we inhabit and intrinsically influence the findings from our qualitative method.

Ethical Considerations and Adverse Impact Statement

This work draws on our real experiences with interacting with various individuals and communities. We recognize that sharing these experiences may raise ethical concerns, including risks to privacy or the possibility of presenting accounts without the perspectives of those involved. To mitigate these risks, we deliberately avoid attributing experiences to identifiable individuals — both ourselves and the people we discuss. Instead, we use generalized descriptions, underspecified language, and a collective narrative voice to minimize the risk that any specific person or community can be identified.

This deliberate underspecification creates a tradeoff with interpretive validity. Many contextual factors that could materially shape experiences of alienation, such as lab type, disciplinary subfield, or the geographic and political context of events referenced in our accounts, are intentionally omitted. The specification of the authors' universities provides limited institutional and disciplinary grounding for our interpretive claims, as a safer contextual proxy. We acknowledge that this constraint limits the degree to which readers can evaluate or situate particular accounts.

This project does not involve human subjects research as defined in 45 CFR 46.102(l) and therefore does not require review by an IRB.

Acknowledgments

We are extremely grateful to the anonymous reviewers and the following individuals for their feedback at various stages of this project: Haley Lepp, Angelina Wang, Sherri Rose, Dan Jurafsky, Jamie Lu, Meera Desai, Evan Dong, Nicky Kriplani, Thalia Zhang, Agata Foryciarz, Zander Majercik, Maria Luiza Rocha Bueno, and Britney Tran. We further appreciate the opportunity to present our work to and receive feedback from the members of the AI, Policy, and Practice Initiative (AIPP) at Cornell University and members of the Health Policy Data Science Lab at Stanford University. Finally, we are indebted to the many people — peers, mentors,

and loved ones — who shaped our thinking on these topics over years of conversation.

References

- AAAI AIES. 2026. 9th AAAI Conference on AI, Ethics, and Society Call for Papers. <https://www.aies-conference.com/2026/call-for-papers/>.
- ACM FAccT. 2026. ACM Conference on Fairness, Accountability, and Transparency (FAccT) 2026 Author Guide. <https://facctconference.org/2026/authorguide.html>.
- Acuna, D. E.; and Liang, L. 2021. Are AI Ethics Conferences Different and More Diverse Compared to Traditional Computer Science Conferences? In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '21, 307–315. New York, NY, USA: Association for Computing Machinery. ISBN 9781450384735.
- Adam, A. 1993. Gendered Knowledge—Epistemology and Artificial Intelligence. *AI & Society*, 7: 311–322.
- Adam, A. 2000. Deleting the Subject: A Feminist Reading of Epistemology in Artificial Intelligence. *Minds and Machines*, 10(2): 231–253.
- Agre, P. E. 1997. Toward a Critical Technical Practice: Lessons Learned in Trying to Reform AI. In Bowker, G. C.; Star, S. L.; Gasser, L.; and Turner, W., eds., *Social Science, Technical Systems, and Cooperative Work: Beyond the Great Divide*. Mahwah, NJ, USA: Lawrence Erlbaum Associates. ISBN 978-0-8058-2403-2.
- Ahmed, S. 2004. *The Cultural Politics of Emotion*. Edinburgh University Press. ISBN 9780748618477.
- Anderson, E. 1995. Feminist Epistemology: An Interpretation and a Defense. *Hypatia*, 10(3): 50–84.
- Anderson, L. 2006. Analytic Autoethnography. *Journal of Contemporary Ethnography*, 35(4): 373–395.
- Bala, P.; Sanches, P.; Cesário, V.; Leão, S.; Rodrigues, C.; Nunes, N. J.; and Nisi, V. 2023. Towards critical heritage in the wild: Analysing discomfort through collaborative autoethnography. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–19.
- Bauer, H. H. 1990. Barriers against interdisciplinarity: Implications for studies of science, technology, and society (STS). *Science, Technology, & Human Values*, 15(1): 105–119.
- Berkes, E.; Marion, M. M.; Milojević, S.; and Weinberg, B. A. 2024. Slow convergence: Career impediments to interdisciplinary biomedical research. *Proceedings of the National Academy of Sciences of the United States of America*, 121(32): e2402646121.
- Beygelzimer, A.; Dauphin, Y.; Liang, P.; and Vaughan, J. W. 2021. Introducing the NeurIPS 2021 Paper Checklist.
- Bietti, E. 2020. From ethics washing to ethics bashing: A view on tech ethics from within moral philosophy. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*, 210–219.
- Birhane, A. 2021. Algorithmic injustice: A relational ethics approach. *Patterns*, 2(2).

- Birhane, A.; Kalluri, P.; Card, D.; Agnew, W.; Dotan, R.; and Bao, M. 2022a. The Values Encoded in Machine Learning Research. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '22, 173–184. New York, NY, USA: Association for Computing Machinery. ISBN 9781450393522.
- Birhane, A.; Ruane, E.; Laurent, T.; S. Brown, M.; Flowers, J.; Ventresque, A.; and L. Dancy, C. 2022b. The forgotten margins of AI ethics. In *Proceedings of the 2022 ACM conference on fairness, accountability, and transparency*, 948–958.
- Boggs, C. 1977. Revolutionary process, political strategy, and the dilemma of power. *Theory and Society*, 4(3): 359–393.
- Bowker, G. C.; and Star, S. L. 2000. *Sorting things out: Classification and its consequences*. MIT press.
- Bundy, J.; Rhodes, V.; Brooks, M.; Brisbane, M.; and Deckard, N. D. 2023. All Things Considered: A Collaborative Critical Autoethnography of Emerging Racialized Scholars. *International Journal of Qualitative Methods*, 22: 16094069231180168.
- Cambo, S. A.; and Gergle, D. 2022. Model Positionality and Computational Reflexivity: Promoting Reflexivity in Data Science. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22. New York, NY, USA: Association for Computing Machinery. ISBN 9781450391573.
- Chan, A. S. 2025. *Predatory Data: Eugenics in Big Tech and Our Fight for an Independent Future*. Oakland, CA: University of California Press. ISBN 9780520402843.
- Chang, H.; Ngunjiri, F.; and Hernandez, K.-A. C. 2016. *Collaborative autoethnography*. Routledge.
- Clough, P. T. 2007. Introduction. In Clough, P. T.; and Halley, J., eds., *The Affective Turn: Theorizing the Social*, 1–33. Durham, NC: Duke University Press.
- Code, L. 1991. *What Can She Know? Feminist Theory and the Construction of Knowledge*. Ithaca, NY: Cornell University Press.
- Colburn, T.; and Shute, G. 2007. Abstraction in Computer Science. *Minds and Machines*, 17(2): 169–184.
- Collins, P. H. 1989. The Social Construction of Black Feminist Thought. *Signs*, 14(4): 745–773.
- Dandelet, S. 2021. Epistemic Coercion. *Ethics*, 131(3): 489–510.
- Daston, L.; and Galison, P. 2007. *Objectivity*. New York: Zone Books.
- Dotson, K. 2011. Tracking Epistemic Violence, Tracking Practices of Silencing. *Hypatia*, 26(2): 236–257.
- Douglas, H. 2000. Inductive risk and values in science. *Philosophy of science*, 67(4): 559–579.
- Ellis, C.; Adams, T. E.; and Bochner, A. P. 2011. Autoethnography: An Overview. *Historical Social Research*, 273–290.
- Ellis, C.; and Bochner, A. P. 2000. Autoethnography, Personal Narrative, Reflexivity: Researcher as Subject. In Denzin, N. K.; and Lincoln, Y. S., eds., *Handbook of Qualitative Research*, 733–768. Thousand Oaks, CA: Sage, 2 edition.
- Fahimi, M.; Russo, M.; Scott, K. M.; Vidal, M.-E.; Berendt, B.; and Kinder-Kurlanda, K. 2024. Articulation Work and Tinkering for Fairness in Machine Learning. *Proc. ACM Hum.-Comput. Interact.*, 8(CSCW2).
- Farmer, P. 2004. *Pathologies of power: Health, human rights, and the new war on the poor*, volume 4. Univ of California Press.
- Freire, P. 1970. *Pedagogy of the Oppressed*. New York: Continuum.
- Fricker, M. 2007. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press.
- Galtung, J. 1969. Violence, peace, and peace research. *Journal of peace research*, 6(3): 167–191.
- Galtung, J. 1990. Cultural Violence. *Journal of Peace Research*, 27(3): 291–305.
- Gebru, T.; and Torres, É. P. 2024. The TESCREAL bundle: Eugenics and the promise of utopia through artificial general intelligence. *First Monday*.
- Green, B. 2021. The contestation of tech ethics: A sociotechnical approach to technology ethics in practice. *Journal of Social Computing*, 2(3): 209–225.
- Hampton, L. M. 2021. Black Feminist Musings on Algorithmic Oppression. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '21, 1. New York, NY, USA: Association for Computing Machinery. ISBN 9781450383097.
- Hanna, A.; Denton, R.; Smart, A.; and Smith-Loud, J. 2020. Towards a critical race methodology in algorithmic fairness. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, FAT* '20, 501–512. New York, NY, USA: Association for Computing Machinery. ISBN 9781450369367.
- Haraway, D. 1988. Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies*, 14(3): 575–599.
- Harding, S. 1992a. After the Neutrality Ideal: Science, Politics, and “Strong Objectivity”. *Social Research*, 59(3): 567–587.
- Harding, S. 1992b. Rethinking Standpoint Epistemology: What is “Strong Objectivity?”. *The Centennial Review*, 36(3): 437–470.
- Hofmann, M.; Kasnitz, D.; Mankoff, J.; and Bennett, C. L. 2020. Living Disability Theory: Reflections on Access, Research, and Design. In *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '20. New York, NY, USA: Association for Computing Machinery. ISBN 9781450371032.
- Husserl, E. 1970. *The Crisis of European Sciences and Transcendental Phenomenology: An Introduction to Phenomenological Philosophy*. Evanston, IL: Northwestern University Press. ISBN 9780810104587.

- Kafar, M.; and Ellis, C. 2014. Autoethnography, storytelling, and life as lived: A conversation between Marcin Kafar and Carolyn Ellis. *Przegląd Socjologii Jakościowej*, 10(3): 124–143.
- Kalluri, P. R.; Agnew, W.; Cheng, M.; Owens, K.; Soldaini, L.; and Birhane, A. 2025. Computer-vision research powers surveillance technology. *Nature*, 643(8070): 73–79.
- Kaltenhauser, A.; Stefanidi, E.; and Schöning, J. 2024. Playing with perspectives and unveiling the autoethnographic kaleidoscope in HCI—A literature review of autoethnographies. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 1–20.
- Keyes, O. 2018. The Misgendering Machines: Trans/HCI Implications of Automatic Gender Recognition. *Proc. ACM Hum.-Comput. Interact.*, 2(CSCW).
- Keyes, O.; Peil, M.; and Barlas, P. 2019. The Misgendering Machines: Trans/HCI Implications of Automatic Gender Recognition. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW): 1–22.
- Khan, R.; Virguez, L.; Paccotacya-Yanque, R.; Mekhael, T.; Munoriyarwa, A.; Salgado, L.; Basu, D.; Mapaling, C.; Perez, N.; Gaudet, Y.; and Larrondo, P. 2025. Whole-Person Education for AI Engineers. In *Proceedings of the Canadian Engineering Education Association (CEEA-ACEG) Conference*. Montreal, Canada.
- Klein, L.; Martin, M.; Brock, A.; Antoniak, M.; Walsh, M.; Johnson, J. M.; Tilton, L.; and Mimno, D. 2025. Provocations from the humanities for generative AI research. *arXiv preprint arXiv:2502.19190*.
- Klumbyte, G.; Draude, C.; and Taylor, A. S. 2022. Critical Tools for Machine Learning: Working with Intersectional Critical Concepts in Machine Learning Systems Design. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '22, 1528–1541. New York, NY, USA: Association for Computing Machinery. ISBN 9781450393522.
- Kuhn, T. S. 1977. *The essential tension: Selected studies in scientific tradition and change*. University of Chicago Press.
- Laufer, B.; Gilbert, T.; and Nissenbaum, H. 2023. Optimization's Neglected Normative Commitments. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '23, 50–63. New York, NY, USA: Association for Computing Machinery. ISBN 9798400701924.
- Leslie, S. W. 1993. *The Cold War and American science: The military-industrial-academic complex at MIT and Stanford*. Columbia University Press.
- Liang, C. A.; Munson, S. A.; and Kientz, J. A. 2021. Embracing Four Tensions in Human-Computer Interaction Research with Marginalized People. *ACM Trans. Comput.-Hum. Interact.*, 28(2).
- Liu, D.; Nanayakkara, P.; Sakha, S. A.; Abuhamad, G.; Blodgett, S. L.; Diakopoulos, N.; Hullman, J. R.; and Eliassi-Rad, T. 2022. Examining Responsibility and Deliberation in AI Impact Statements and Ethics Reviews. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '22, 424–435. New York, NY, USA: Association for Computing Machinery. ISBN 9781450392471.
- Longino, H. E. 2020. Science as social knowledge: Values and objectivity in scientific inquiry.
- Lorde, A. 1981. The Uses of Anger: Women Responding to Racism. In *Sister Outsider: Essays and Speeches*. Freedom, CA: Crossing Press.
- Lugones, M.; and Spelman, E. V. 1983. Have We Got a Theory for You! Feminist Theory, Cultural Imperialism and the Demand for “The Woman’s Voice”. *Women’s Studies International Forum*, 6(6): 573–581.
- Mäkinen, E. I.; Evans, E. D.; and McFarland, D. A. 2025. Interdisciplinary Research, Tenure Review, and Guardians of the Disciplinary Order. *The Journal of Higher Education*, 96(1): 54–81.
- Malazita, J. W.; and Resetar, K. 2019. Infrastructures of abstraction: How computer science education produces anti-political subjects. *Digital Creativity*, 30(4): 300–312.
- Malik, M.; and Malik, M. M. 2022. Critical technical awakenings. *Journal of Social Computing*, 2(4): 365–384.
- Marx, K. 1978. Estranged Labour. In Tucker, R. C., ed., *The Marx–Engels Reader*, 66–125. New York: W. W. Norton, 2 edition. Originally written in 1844.
- Metcalfe, J.; Moss, E.; et al. 2019. Owning ethics: Corporate logics, silicon valley, and the institutionalization of ethics. *Social Research: An International Quarterly*, 86(2): 449–476.
- Mohamed, S.; Png, M.-T.; and Isaac, W. 2020. Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence. *Philosophy & Technology*, 33: 659–684.
- NAACL 2022 Organizing Committee. 2021. New Checklist for Responsible NLP Research.
- Nagel, T. 1986. *The View from Nowhere*. New York: Oxford University Press.
- Nel, N. 2018. Relational collaborative autoethnography: Post-doctoral fellowship in South Africa. *Participatory Educational Research*, 5(1): 60–73.
- Ngunjiri, F. W.; Hernandez, K.-A. C.; and Chang, H. 2010. Living autoethnography: Connecting life and research. *Journal of research practice*, 6(1): E1–E1.
- Nixon, R. 2011. *Slow Violence and the Environmentalism of the Poor*. Harvard University Press.
- Noel, T. K.; Minematsu, A.; and Bosca, N. 2023. Collective Autoethnography as a Transformative Narrative Methodology. *International Journal of Qualitative Methods*, 22: 16094069231203944.
- Ogbonnaya-Ogburu, I. F.; Smith, A. D.; To, A.; and Toyama, K. 2020. Critical Race Theory for HCI. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, 1–16. New York, NY, USA: Association for Computing Machinery. ISBN 9781450367080.
- Olteanu, A.; Ekstrand, M.; Castillo, C.; and Suh, J. 2023. Responsible AI Research Needs Impact Statements Too.

- Peterson, T. L.; Ferreira, R.; and Vardi, M. Y. 2023. Abstracted power and responsibility in computer science ethics education. *IEEE Transactions on Technology and Society*, 4(1): 96–102.
- Raji, I. D.; Scheuerman, M. K.; and Amironesei, R. 2021. You can't sit with us: exclusionary pedagogy in AI ethics education. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, 515–525.
- Russo, M.; Jorgensen, M.; Scott, K. M.; Xu, W.; Nguyen, D. H.; Finocchiaro, J.; and Olckers, M. 2024. Bridging Research and Practice Through Conversation: Reflecting on Our Experience. In *Proceedings of the 4th ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, EAAMO '24. New York, NY, USA: Association for Computing Machinery. ISBN 9798400712227.
- Rutter, N.; Hasan, E.; Pilson, A.; and Yeo, E. 2023. "It's the End of the PhD as We Know it, and We Feel Fine... Because Everything Is Fucked Anyway": Utilizing Feminist Collaborative Autoethnography to Navigate Global Crises. *International Journal of Qualitative Methods*, 22: 16094069211019595.
- Scheper-Hughes, N.; and Bourgois, P. 2004. Introduction: Making Sense of Violence. In Scheper-Hughes, N.; and Bourgois, P., eds., *Violence in War and Peace: An Anthology*, 1–27. Oxford: Blackwell.
- Scheuerman, M. K.; Branham, S. M.; and Hamidi, F. 2018. Safe Spaces and Safe Places: Unpacking Technology-Mediated Experiences of Safety and Harm with Transgender People. *Proc. ACM Hum.-Comput. Interact.*, 2(CSCW).
- Schroeder, H.; Pareek, A.; and Barocas, S. 2025. Disclosure without Engagement: An Empirical Review of Positionality Statements at FAccT. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '25, 1195–1210. New York, NY, USA: Association for Computing Machinery. ISBN 9798400714825.
- Selbst, A. D.; Boyd, D.; Friedler, S. A.; Venkatasubramanian, S.; and Vertesi, J. 2019. Fairness and Abstraction in Sociotechnical Systems. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 59–68.
- Selinger, E.; and Durant, D. 2022. Amazon's ring: Surveillance as a slippery slope service. *Science as culture*, 31(1): 92–106.
- Shelby, R.; Rismani, S.; Henne, K.; Moon, A.; Rostamzadeh, N.; Nicholas, P.; Yilla-Akbari, N.; Gallegos, J.; Smart, A.; Garcia, E.; and Virk, G. 2023. Sociotechnical Harms of Algorithmic Systems: Scoping a Taxonomy for Harm Reduction. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '23, 723–741. New York, NY, USA: Association for Computing Machinery. ISBN 9798400702310.
- Shew, A. 2023. *Against Technoableism: Rethinking Who Needs Improvement*. New York, NY: W. W. Norton & Company.
- Singh, A.; Dechant, M. J.; Patel, D.; Soubutts, E.; Barbareschi, G.; Ayobi, A.; and Newhouse, N. 2025. Exploring Positionality in HCI: Perspectives, Trends, and Challenges. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, CHI '25. New York, NY, USA: Association for Computing Machinery. ISBN 9798400713941.
- Spivak, G. C. 1988. Can the Subaltern Speak? In Nelson, C.; and Grossberg, L., eds., *Marxism and the Interpretation of Culture*, 271–313. Urbana: University of Illinois Press.
- Stahlke Wall, S. 2016. Toward a Moderate Autoethnography. *Qualitative Research in Organizations and Management: An International Journal*, 11(1): 6–19.
- Star, S. L.; and Bowker, G. C. 2007. Enacting silence: Residual categories as a challenge for ethics, information systems, and communication. *Ethics and Information Technology*, 9(4): 273–280.
- Stark, L. 2019. Facial recognition is the plutonium of AI. *XRDS: Crossroads, The ACM Magazine for Students*, 25(3): 50–55.
- Suchman, L. 2015. Situational Awareness: Deadly Bioconvergence at the Boundaries of Bodies and Machines. *Medi-aTropes*, 5(1).
- Suchman, L.; Gururaja, S.; and Widder, D. G. 2025. AI interdisciplinarity as critical technical practice. *Cambridge Forum on AI: Culture and Society*, 1: e2.
- Waelen, R. A. 2024. The ethics of computer vision: An overview in terms of power. *AI and Ethics*, 4(2): 353–362.
- Wang, A. 2025. Identities are not Interchangeable: The Problem of Overgeneralization in Fair Machine Learning. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '25, 485–497. New York, NY, USA: Association for Computing Machinery. ISBN 9798400714825.
- Widder, D. G. 2024. Epistemic Power in AI Ethics Labor: Legitimizing Located Complaints. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '24, 1295–1304. New York, NY, USA: Association for Computing Machinery. ISBN 9798400704505.
- Williams, R. M. 2019. Metaeugenics and Metaresistance: From Manufacturing the 'Includeable Body' to Walking Away from the Broom Closet. *Canadian Journal of Children's Rights / Revue Canadienne des Droits des Enfants*, 6(1): 60–77.
- Williams, R. M. 2025. *Disabling Intelligences: Legacies of Eugenics and How We are Wrong about AI*. Cham: Springer Nature Switzerland. ISBN 978-3-032-02665-1.
- Ymous, A.; Spiel, K.; Keyes, O.; Williams, R. M.; Good, J.; Hornecker, E.; and Bennett, C. L. 2020. "I am just terrified of my future" — Epistemic Violence in Disability Related Technology Research. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI EA '20, 1–16. New York, NY, USA: Association for Computing Machinery. ISBN 9781450368193.
- Young, M.; Katell, M.; and Krafft, P. 2022. Confronting power and corporate capture at the FAccT Conference. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 1375–1386.

Zheng, X.; Peng, A.; Hong, X.; Sugimoto, C. R.; and Ni, C.
2025. Interdisciplinary PhDs face barriers to top university
placement within their disciplines. [arXiv:2503.21912](#).