

Moral Missions: Sustaining Responsible Practice by Surfacing Practitioner Moral Decision-Making Procedures

Teanna Barrett¹, B. Biira², Jainaba Jawara³, Andrew Shaw^{1,4}, Ziwei Dong⁵, Chinasa T. Okolo⁶, Seyi Olojo⁷, Keerthana Kompella¹, Khadija Saho², Amy X. Zhang¹, Leilani Battle¹

¹Paul G. Allen School of Computer Science & Engineering, University of Washington

²Information School, University of Washington

³University of Maryland College of Information

⁴Department of Computer Science, Cornell University

⁵Microsoft

⁶Technecultura

⁷School of Information, University of California Berkeley

tb23@cs.washington.edu

Abstract

A growing ecosystem of techniques, toolkits, and guidelines has been developed to help data scientists consider the social implications of data-driven technologies. However, prior literature highlights that even when this ecosystem of techniques is provided to professional data scientists, they still struggle to consistently adopt a responsible data science practice. We posit that the key to sustained responsible data science practice is to approach it as a “*moral mission*”: a conviction-driven technical practice that seeks to transform social conditions by any degree possible. In this paper, we present a semi-structured interview study with 15 responsible data scientists and AI practitioners to understand the moral decision-making procedures they use to articulate and actualize their moral missions. Through a phenomenological analysis of our participants’ accounts, we find participants engage in embodied introspection, circumvent institutional expectations, and center relationality throughout their moral missions. We also present how our participants engage in similar processes to contend with generative AI (GenAI) in their responsible practice. We conclude by calling for subversive data science communities and identifying sociotechnical design implications to better support sustainable responsible data science practice.

Introduction

Despite a growing ecosystem of frameworks, toolkits, and guidelines for responsible data science (RDS), practitioners struggle to meaningfully integrate these resources into everyday practice (Wong, Madaio, and Merrill 2023). One obstacle to responsible practice is the deprioritization of RDS techniques within organizations, particularly amid shifting political and institutional interests that impact sustained attention to artificial intelligence (AI) ethics. Consequently, upholding and actualizing RDS is often left to individual practitioners (Metcalf, Moss et al. 2019). Prior work has largely focused on outlining how practitioners should practice RDS; less scholarship pays attention to how practitioners make moral decisions in situ (Madaio et al. 2020, 2022).

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

Moral decision-making procedures capture the experience of engaging with a set of principles, codes, or values to guide one’s actions (Shipp, LaLonde, and Martinez 2023; Smith 2018; Preti 2018). From a pragmatist perspective, these processes are rarely idealized applications of right or wrong but complex assessments of the situation, available decisions, and the applicability of one’s convictions (JafariNaimi, Nathan, and Hargraves 2015). Individual moral decision-making procedures that mediate between values and technical action remain underexplored (Wang, Datta, and Dickerson 2024; Ali et al. 2023). This gap is critical as responsible practice is not solely determined by the availability of tools, but by the situated judgments, constraints, and value negotiations that shape their use within organizational contexts.

We argue that supporting RDS requires understanding moral decision-making procedures as an explicit part of RDS practice. To this end, we focus our study of RDS practice on capturing defining decisions that are often overlooked and implicit strategies that sustain responsible practice.

We present a qualitative phenomenological study of fifteen self-identified responsible data scientists to surface the moral reasoning procedures that guide applications of responsible techniques under real-world constraints. We draw on interpretive phenomenological analysis (IPA) to investigate the lived moral reasoning that guides how practitioners select, adapt, and deploy RDS techniques.

Our study is guided by the following research questions:

1. **How do data scientists committed to building responsible data-driven technology identify their personal values?**
2. **How do data scientists committed to building responsible data-driven technology adapt and integrate responsible data science techniques into their work?**

From our participants’ accounts of practicing RDS, we conceptualize the phenomenon as a “*moral mission*”. We define moral missions as conviction-driven technical practices that seek to transform social conditions by any de-

gree possible. While embarking on their moral missions, we find our participants **1)** navigate through deepening levels of embodied introspection to define their professional convictions, **2)** pragmatically weigh the tradeoffs of working within or circumventing institutional expectations to guide their technical approach, and **3)** look outside of themselves to nurture their practice. Our participants also demonstrate moral mission decision-making procedures when considering the use or refusal of Generative AI (GenAI) within existing responsible techniques. By engaging with Moten and Harney (2013)’s theory of the *undercommons*, we recognize the expansive potential of approaching RDS practice as a moral mission and informing design directions for RDS tools. Overall, our work contributes to AI ethics scholarship by centering the internal world of professional data scientists to surface moral decision-making procedures RDS tools should be built to support.

Related Work

Values in Design and Responsible Data Science

Although scholars have long recognized that technology can shape and be shaped by human values (Winner 1980), theories on how designers of technologies can and should actualize values in their work are ever-evolving. One of the first influential accounts of values in technological design is Value Sensitive Design (VSD), which defines a value as “what is important to people in their lives, with a focus on ethics and morality” (Friedman et al. 2006). VSD employs iterative methodologies by integrating conceptual, empirical, and technical investigations to help designers navigate the relationship between values and technology.

However, the scope and nature of moral values remain an ongoing topic of discussion within VSD work: while a universalistic approach to human values risks imposing colonialist constructs onto local contexts, a relativistic approach risks condoning any practice as a valid expression of culture. Whereas earlier work from Friedman and Kahn (2002) endorses a conception of values as universal concepts that can have different instantiations in particular cultures, later work from Borning and Muller (2012) rejects the universalist conception of values in favor of a pluralistic one. They also recognize that non-exhaustive heuristic lists of values (like those put forward by Friedman et al. (2006)) may nonetheless be valuable in practice.

The study of values in design is particularly relevant to RDS, a domain in which scholars have long highlighted the importance of situated and open-ended value inquiry (Morley et al. 2021). D’Ignazio and Klein (2020), for example, draw out seven principles of Data Feminism from feminist use cases of data science, including challenging power, embracing pluralism, and considering context. Other situated frameworks draw from feminist theories of justice (Ajmani et al. 2024), decolonial theories (Mohamed, Png, and Isaac 2020), and African philosophies (Barrett et al. 2025) to highlight how legacies of power and collective experiences shape the curation and interpretations of RDS principles.

In addition to these frameworks, other scholars have also taken more universalist approaches to survey RDS values.

Representative examples include AI4thePeople, a framework with five values created by Floridi et al. (2018). Shen et al. (2025) employ a set of 56 human values drawn from Schwartz (1994) to measure value alignment between humans and LLMs across different cultures and scenarios. We see the rich theoretical scholarship of value-laden design and technical practice as powerful probes for reflection. Hence, we draw values from both relativist and universalist value frameworks mentioned above when designing value cards used in our interviews.

VSD has also been critiqued for following what JafariNaimi, Nathan, and Hargraves (2015) call an “identify/apply” logic, which assumes moral problem-solving is conceptualized as a process of faithfully applying values that are already known to stakeholders. However, JafariNaimi, Nathan, and Hargraves (2015) object that this “common-place understanding of values in VSD as *what a person or group considers important in life* [...] emphasizes the identification of values rather than their service in design situations.” Although VSD acknowledges that conceptual investigation should be integrated with empirical and technical investigation, treating *values as hypotheses* (coined by JafariNaimi, Nathan, and Hargraves (2015)) means that values are most comprehensively investigated within introspective, storied inquiries of practitioners’ design experiences. Therefore, we center the storied inquiries of responsible data scientists to better understand the relationship between RDS values and practice.

Responsible Data Science Techniques

Researchers have developed numerous RDS techniques to address ethical concerns in AI development. Documentation tools have standardized how practitioners document metadata, data collection, methods, considerations, and intended use. Examples include Datasheets for Datasets (Gebru et al. 2021) and Model Cards (Mitchell et al. 2019). Fairness libraries such as Bellamy et al. (2018)’s AI FAIRNESS 360 (AIF360) and Bird et al. (2022)’s Fairlearn are open-source toolkits to help practitioners detect, measure, and mitigate bias in their models. On the other hand, fairness libraries embed specific mathematical definitions of fairness that might not align with domain understanding of justice (Hertweck, Heitz, and Loi 2024). Explainability techniques aim to make model behavior more understandable by targeting transparency (Lipton 2017), though critics argue interpretability does not guarantee stakeholders can act on explanations or that transparency alone shifts power dynamics (Selbst et al. 2019). Practitioners also regularly use less formalized data cleaning, demographic balancing, and anonymization techniques (Hsu et al. 2021). All of these techniques can degrade into performances of ethics that fit within organizational logics and misdirect efforts for meaningful change (Metcalf, Moss et al. 2019).

Interactive tools developed in HCI research aim to make ethical deliberation more accessible and participatory. Cards and gamification have been used as approaches to navigating ethics in HCI research (Friedman et al. 2006; Widder et al. 2024). An example is Value Cards (Shen et al. 2021), which helps teams navigate through conflicting values during the

design process. Bitácora is built for nonprofits using social media data; it recognizes that organizational context matters for what counts as responsible (Alvarado Garcia et al. 2024). Checklists also aim to provide practitioners with clear and actionable guidance (Madaio et al. 2020). While these approaches give structured time to explore ethical dilemmas, they all require collective time (Lin, Wang, and Yueh 2022; Murray-Rust et al. 2023).

A further challenge for RDS tools pointed out by Wong, Madaio, and Merrill (2023) is that most tools are siloed, addressing isolated problems rather than integrating into actual workflows, leaving practitioners to figure out which tools matter and whether they can use them at work. Crucially, these tools tell practitioners what to evaluate but offer little guidance on how to understand their own positionality when taking value-laden sociotechnical actions. The rise of GenAI also exposes limitations in existing RDS techniques. Wolfe and Mitra (2024) show that GenAI introduces new problems, including attribution, consent, and environmental impact, that existing tools were not designed to address.

Evaluating Responsible Data Science Practices

Prior work in operationalizing RDS has largely focused on proposing and evaluating lifecycle-oriented frameworks, capturing the challenges of toolkit adoption, and surfacing workflows and perceptions from practitioners (Zhang, Muller, and Wang 2020; Heger et al. 2025; Krijger et al. 2023). RDS frameworks outline the moral concerns that arise across the entire data science process, from problem formulation and data collection to modeling, evaluation, deployment, and communication (Saltz and Dewar 2019; Stoudt et al. 2024; Urovi et al. 2024; Boenig-Liptsin, Tanweer, and Edmundson 2022). To understand the technical needs and feasibility of RDS, prior HCI scholarship has explored the efficacy of existing RDS toolkits (Deng et al. 2022). Works from Varanasi and Goyal (2023); Bhattacharjee and Ribes (2025) acknowledge how institutional structures impact the ability for AI and ML practitioners to uphold responsible AI values. The challenge grows as corporations increase adoption of GenAI in engineering workflows.

Emerging work further aims to surface the introspective beliefs, deliberations, and strategies of RDS to inform a more holistic and generative understanding of RDS practice to better support it (Bennett 2025; Madaio et al. 2024). Dong et al. (2026) offers a useful lens for addressing this gap by reframing RDS as a problem of practitioner behavior change. Rather than evaluating responsibility only through models, datasets, metrics, or organizational policies, responsible outcomes also depend on the everyday behaviors and decisions of data scientists (Dong et al. 2025). We build on these prior works by emphasizing that evaluating RDS ultimately requires understanding how practitioners repeatedly enact those values in their concrete decisions.

Methods

Our study was approved by our institutional review board and pre-registered with the Open Science Foundation.¹

¹<https://osf.io/g56fd>

Moral Missions Interview Study Protocol

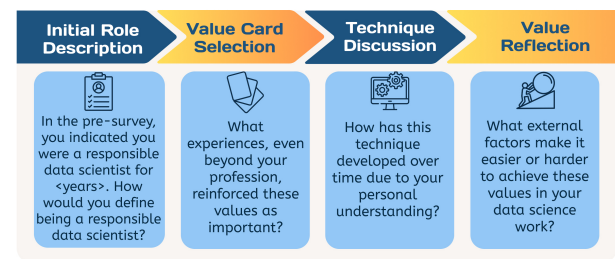


Figure 1: Semi-structured interview used to elicit participants' RDS values, selected techniques, and reflections on the conditions that shape value enactment in practice. Example questions from the interview protocol are featured at each stage.

Participants

For this study, we aimed to recruit participants who identify as RDS scientists or AI practitioners in industry, government, and non-profit sectors. We recruited participants through RDS Slack channels, LinkedIn posts, local RDS meet-ups, and word of mouth throughout the research team's network. All participants completed a pre-survey to detail their data science experience and focus, identify an RDS technique to discuss in the interview, provide optional demographic information, and finally confirm their consent and accessibility accommodations (breaks, captioning, access to questions beforehand). In the end, we recruited 15 participants from diverse industries, experience levels, and data science roles. The interviews on average lasted 82 minutes, and participants were compensated \$25 for their time. An overview of our participant demographics is presented in Table 1.

Interview Study Design

To gain insight into the lived experiences of data scientists using RDS techniques, we designed a semi-structured interview with an interactive visual probe. The visual probe was a workspace in which participants could annotate, interact with pre-made value and technique cards, and add online media. The interview protocol was split into four phases: initial role description, value card selection, technique discussion, and finally, a value-grounded reflection of their practice (refer to Figure 1).

Data Science Role Description. In the initial role description phase, participants described their current role as a data scientist. We asked participants to describe the data they work with, the products, systems, or initiatives they contribute to, and the colleagues, communities, or stakeholders they interact with.

Data Science Value Card Selection. After providing an overview of their work, we introduced the participant to the interactive visual probe workspace. Participants were given time to familiarize themselves with the interface and then asked to select three to five value cards they personally deemed important to their work as a responsible data scientist. The value card form and selection activity are in-

Alias (<i>pronouns</i>)	Role	Industry	Data Science Experience (Yrs)	Country of Residence	Ethnicity	RDS Technique
P1 (<i>they/them</i>)	Battery Scientist	Corporate	8	X	Asian	Up & Down Sampling
P2 (<i>she/her</i>)	Data Scientist	Government	5	United States	European or White	Pre-mortem Documentation
P3 (<i>X</i>)	Data Scientist	Government	10	X	X	Removing PII
P4 (<i>she/her</i>)	Product Manager	Corporate	12	United States	European or White	Model Documentation
P5 (<i>he/him</i>)	AI Practitioner	Non-Profit	3	Nigeria	Black	Human-in-the-loop Advocacy
P6 (<i>she/her</i>)	Machine Learning Engineer	Corporate	3	Canada	Black	Model Cards (Mitchell et al. 2019)
P7 (<i>he/him</i>)	Data Scientist	Government	3	United States	Asian	Datasheets for Datasets (Gebru et al. 2021)
P8 (<i>X</i>)	Research Analyst	Non-Profit	3	United States	Black	Removing PII
P9 (<i>she/her</i>)	Developer	Non-Profit	5	Canada	European or White	Serverless Data Governance
P10 (<i>she/her</i>)	Modeler	Non-Profit	5	Nigeria	Black	Model Calibration Documentation
P11 (<i>she/her</i>)	Data Scientist	Corporate	6	X	Asian	SHAP (Lundberg and Lee 2017)
P12 (<i>he/him</i>)	AI/ML Engineer	Non-Profit	4	Nigeria	Black	Subject Matter Expert Consultation
P13 (<i>he/him</i>)	Principal Data Scientist	Corporate	18	United States	European or White	Distributive Visualizations
P14 (<i>he/him</i>)	AI/ML Engineer	Corporate	5	Nigeria	Black	Reducing API Use
P15 (<i>he/him</i>)	Marketing Data Engineer	Corporate	6	Mexico	X	Encrypting PII

Table 1: A table summarizing characteristics of each 15 participants, including pronouns, data science role, industry, years of data science experience, ethnicity, country of residence, and the responsible data science technique they discussed in their interview. Note: optional information participants chose not to disclose indicated with an X.

spired by VSD Envisioning Cards (Yoo et al. 2022) and AI ethics education value cards (Shen et al. 2021). The provided value cards were curated from universal design value research and ethics principles informed by specific epistemic communities: Data Feminism (Klein and D’Ignazio 2024), AI4thePeople (Floridi et al. 2018), African Data Ethics (Barrett et al. 2025), and Disability Justice principles (Berne et al. 2018).

Responsible Data Science Technique Discussion. After selecting of value cards, we moved to the RDS technique discussion. We defined RDS techniques as technical practices within the data science workflow to improve, consider, mitigate, or evaluate social implications of data-driven technology. Participants were asked to identify a technique to discuss for the duration of the interview. We also provided technique cards to inspire participants based on reviews of RDS tools (Dong et al. 2025; Kuehnert et al. 2025). In the discussion of the technique, we asked them to identify value cards relevant to their technique of choice, how they learned about the technique, use cases, and process to implement the technique.

Value Reflection. Finally, we asked the participants to reflect on their technique through the lens of their values. Participants discussed how their values were actualized in practice, what external factors influenced this actualization, and how their larger professional community should engage with RDS.

Data Analysis

The data were primarily collected through transcribed and recorded interviews. Additional data were collected through the pre-survey, interview notes, and the visual probe artifact. All of the participants’ anonymized and cleaned data were categorized in folders with de-identified participant ID labels. Participants with identifiability concerns were offered the opportunity to review relevant data (i.e., recordings and transcripts) and select items for anonymization or removal. Participants were also informed that they could exit the study at any point, even upon study completion, and could ask for relevant collected data to be excluded.

Qualitative Analysis Methodology. We developed an analysis plan with a foundation in interpretative phenomenological analysis (IPA) (Dahlberg and Dahlberg 2020). Phenomenological qualitative analysis aims to explore lived experiences by centering the individual’s ability to articulate a comprehensive account of a meaningful activity (Eatough and Smith 2017). Phenomenological analysis has been deployed in HCI studies to gain a deep understanding of user needs, existing technology use, and design implications (Linder and Arvola 2017; Kallia 2025; Midha, Wilson, and Sharples 2022; Horvath and Raptis 2025; Fortin et al. 2013; Rizvi et al. 2025). Its strengths are especially apparent in capturing novel, introspective activities such as moral decision-making in RDS practice. Each participant’s understanding of their personal and social world grounds analysis and interpretation in individual nuance often smoothed over in thematic analysis (Midha, Wilson, and Sharples 2022).

Analysis Approach. A subset of the research team (two interview facilitators and a non-interviewing researcher) performed the qualitative analysis using Atlas.ti. Data were evenly distributed among the team. We conducted three rounds of analysis to construct a phenomenological account of each participant’s process of integrating and adapting their example of an RDS science technique. First, the team reviewed each participant’s data sources multiple times to facilitate familiarization. During this phase, one-page documents were created for each participant to capture potential design considerations arising from their experiences. The team then coded transcripts and visual probe artifacts to identify aspects of interest. We developed narrative accounts for each interview to document our interpretation of how participants’ experiences addressed our research questions, and we met as a team to review them collectively, apply theoretical concepts, and reach interpretive consensus. Finally, we drew on the resulting codes, narratives, design considerations, and theoretical touchpoints to generate themes, an overarching conceptual model, and design implications (Naeem et al. 2023).

Findings

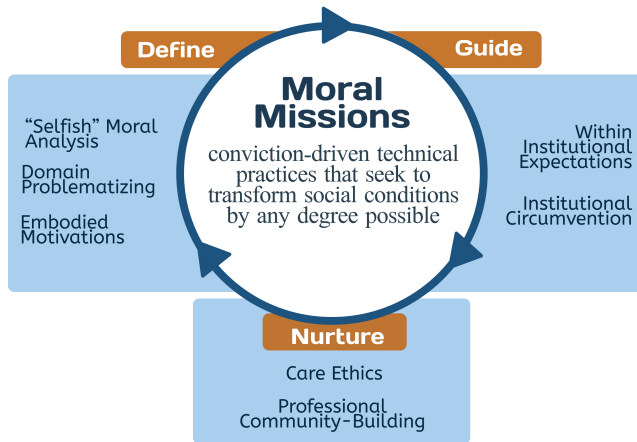


Figure 2: Conceptual model of moral missions as an iterative process of defining moral commitments, guiding RDS practice under institutional constraints, and nurturing commitments through relational and professional communities.

We analyze our fifteen participants' moral decision-making procedures through discussions of personal data science values (**RQ1**) and accounts of implementing RDS techniques (**RQ2**). Our analysis demonstrates the phenomena we call *moral missions* are: 1) defined through embodied introspection, 2) guides for innovative circumventions of institutional rigidity, and 3) nurtured by relationality beyond the workplace. To close, we highlight a case study of these concepts in action with the emergence of GenAI.

RQ1: How do data scientists committed to building responsible data-driven technology identify their personal values?

Defining Moral Missions: Embracing Embodied Introspection for Conviction Development

At first, many participants were uncertain or hesitant about articulating their moral priorities. Once they engaged with our interview probes, participants revealed layers of moral reasoning that ultimately led to intimate and emotionally grounded motivations for their practice.

“Selfish” Moral Analysis We asked participants to select a subset of value cards they deem most important to their responsible practice. Several participants explained they often don't have the space within their workplace to explicitly discuss their values (P1, P2, P6, P10, P15). Therefore, many participants exhibited reservations about exercising moral agency. For example, P2 qualified their creation of the value *Collective Well-Being* with selfishness:

I think I tend to be pretty empathetic as a person. So it's almost selfish; the sense that other people are doing okay around me [...] is very comforting. I mean, that's sort of the selfish angle; I'm comforted by other people.

A subset of participants who lead responsible data-driven organizations were more assured of their moral agency and readily bypassed our proposed values to present their own (P9 and P13). For instance, P13 created a value card for *Epistemic Responsibility*:

Epistemic responsibility is one view of an accurate, ethical way to show data. [...] I think our greater responsibility is not saying accurate things; it's promoting accurate beliefs. This is maybe related to consequentialism, but thinking about the effects, how things are perceived, and taking responsibility for a few more steps down the chain of communication.

Domain Problematicizing Across all participants, values were jumping-off points to problematize and subsequently construct moral analyses of their data science domain. For example, P1 models electric vehicle (EV) battery performance based on battery data across EV manufacturers. In selecting *Consider Power*, P1 articulated how American companies and their powerful supporters determine data access and, in turn, shape domestic EV purchasing decisions in their favor.

Teslas are the majority of used electric vehicles out there, and a lot of that has to do with the amount of influence and power that company has been in successfully selling their products. There are a lot of very well-made Chinese models that do not exist in the US because of import rules. There are a lot of [original equipment manufacturers] (OEMs), manufacturers, core sellers in the US that do not want Chinese brands into the US. So that influences the amount of data that we have and what vehicles we can serve.

While considering *Freedom from Bias*, P15 realized his work in targeted marketing contradicted his moral inclination.

On the marketing side, you just want to create the audiences, so targeting, exclusiveness, or going in that direction would be what people are looking for. [...] I mean, at the point of the arrow, the data references should be [...] clean and free of bias [...] I mean, I will not [select 'Freedom from Bias'] but I just want to bring this idea out of my head.

Later in the interview, P15 was asked to consider what it would look like to practice targeted marketing free from bias. He acknowledged he had not considered this tension before, but imagined regulations on advertising frequency were in the right direction. Across these instances, problematizing through values surfaced implications of participants' work and compelled them to take a position.

Embodied Motivations to Pursue Data Science As the discussion of values continued, participants moved beyond abstract morality or domain knowledge to share the intimate experiences that established their emotional connection to RDS (Klein and D'Ignazio 2024). For example, P8 and P12 returned to their hometowns. P8 grew up without many resources and did not appreciate the “savior complex” non-African NGOs adopted to fundraise for policy interventions.

He assumed his role as a policy data analyst to inform narratives that respect the dignity of underresourced communities. P12 was born in North-Central Nigeria and grew up watching people around him struggle. When he found out about multilingual language models, he saw it as a powerful means to improve his community's conditions.

I seek to contribute to the lives of people, probably starting with just one person. Someone who is smart but struggles, whose parents cannot cater for them. I'm committed to helping improve the quality of life of people [...] The whole idea is how can locals in the interior parts of the continent access the same technology that people in the diaspora and in the urban areas enjoy.

P2, P4, and P11 were drawn to RDS by their personal experiences as women. P11, particularly as a woman of color, observed how systems—AI included—can treat people unfairly based on identity:

Having grown up as a woman of color, [I have seen] when these biases and stereotypes people have are enforced in different situations and how harmful that can be, and how much upfront unfairness a lot of people suffer. So this has very much informed my work. With my personal lived experiences, AI and machine learning [should] not include these factors at all, instead of saying: oh no, we are using gender, but we are not really using it to discriminate. How do you know? How much control do you have over an AI or ML model that you can for sure say that?

Most participants cited formal and informal training as formative to internalizing the importance of their responsible practice. P14 shared how a single conversation with a colleague shifted his perspective on third-party API use:

Two years ago, I was working on a project with a colleague who is also an AI engineer. One of the discussions we were having about a project, he spoke about the fact that [...] users would have to enter information about their therapies and things like that, which is quite private. He highlighted the fact that we would have to reduce third-party API use as much as possible, because these are private information and the third-party developers of these APIs can't have access to this information [...] I'd always known that third-party APIs would more likely have access to the information we bring in, but I've mostly not cared. If I get my work done and it does the job, just go ahead and use it. But then having that conversation opened my eyes to seeing that we have to be more careful.

RQ2: How do data scientists committed to building responsible data-driven technology adapt and integrate responsible data science techniques into their work?

Moral Missions as Guides: Navigating Institutional Expectations

Participants' moral missions guided them to make the most responsible decisions within their institution's expectations

or circumvent institutional logics to maintain their convictions.

Working within Institutional Expectations Participants ran into challenges at several levels. At the institutional level, responsible practices often lose out to business priorities. For example, P4 described how model and data documentation were never prioritized by the leadership:

From a business perspective, it's not ever going to be the highest priority item to solve. The bottom line is profit is the number one driver until someone breaks something. Then it's the highest priority to solve, but it's usually not thinking long-term.

Working within these institutions often involved acknowledging the current limitations of the data or systems the practitioner was working with. Participants like P1, P4, P10, and P11 did their best to communicate sociotechnical limitations to their non-technical colleagues to exercise a form of moral agency in their work. P1 described flagging limitations to product and marketing colleagues before decisions were made:

I try to keep in mind, like, oh, these are the biases that we're dealing with. And these are the limitations, so maybe don't trust this result as much. Even though the stat says there's a 95% probability that these two groups are different by X amounts. That's without the context of how I trust the data. So then, with my input, it's like, oh, maybe you don't make a decision based on this.

When evaluating the efficacy of working within their explicit institutional role, participants who felt limited by their institution speculatively turned to policy change and regulations to expand their opportunities for responsible practice (P1, P5, P15). These participants' hope in policy was confirmed by P11, who was not worried about RDS losing priority within her institution, even with the rise of GenAI:

When you have regulatory oversight, they cannot [de-prioritize explainability], because it's not dependent on them. Regulators are probably the biggest engine of change I have seen with respect to moving towards explainability.

Participants also described negotiating with funders, clients, policymakers, and government officials to reconcile differing priorities and definitions of responsible practice. P10 models malaria public health interventions to predict which local government areas would benefit the most. She describes how she presents her model to program managers as a support tool for equitable decision-making:

We always let the program manager [know that] modeling is a decision support tool. These are things which help them to at least have some evidence to objectively make decisions. Maybe [previously, they] based their decisions on [their] state of origin. [...] If you channel the resources you have [for] maximum impact, incrementally the margin of reduction [of malaria transmission] will be more.

Strategies to Circumvent Institutional Expectations

For some participants, institutional pushback expanded their moral missions to devise innovative sociotechnical strategies (P2, P12, P13). P9 held her team's position even when it cost her a client:

There was a question from the government about where our servers were located to protect the data privacy [...] it took us a couple of weeks to realize we didn't have that much data, and we didn't need to store data. We really didn't need to. So then we'd gone back to the government and said, 'we don't have any servers; we're not getting any.' And they wouldn't work with us anymore, because the government had a rule that said that the data had to be safeguarded in your own server, in your own data container [...] We were sticking to our rules, our values about what we were going to do.

P13 completed a project “out of spite” to prove his novel, equity-focused technique can rival traditional visualization approaches.

I had another client that was actually social advocacy focused, and they were like, 'no, no way, we're gonna do the traditional thing, that's what's gonna sell it to funders.' That was weird and frustrating, but it also kind of triggered this; I'm pretty sure I can prove you wrong. I want to prove that this is real. From the startup world, I had a sense of how to run an A-B test and all that. So, [I] just started testing some of the ideas to see [whether] certain ways of presenting disparity data lead to stereotypes or lead to victim blaming. It was a thing that was speculated online, but there wasn't anything empirical on it. So, mainly out of spite, I ran those experiments to just see what would happen. It turns out that, yeah, it does have an effect and got results on that.

In more hostile institutional environments, such as P2's, surface-level compliance created a cover for technical resistance:

One of the ways that I've seen models around by other people in the civil service has been continuing to do the work, but just calling it something else [...] There's a lot of integrity that I saw of people just, like, well, this is the type of data that we know people find useful, so we're gonna keep doing the work. As the [executive orders] (EOs) were rolling out, and they're like, can't talk about women— okay, fine, just take the word women down [from the documentation].

Nurturing Moral Missions: Relationality and Collective Decision-Making

Regardless of the decisions made by participants, they knew they needed to consider relationality in their technical actions to take their moral missions to the next level (Birhane 2021). Throughout their accounts, participants made a point to acknowledge end users, impacted communities, or supportive professional communities.

Care Ethics For many of our participants, the most crucial decision-making parameter was the impact on the end user or data subject population. Several participants' practices resonated with theories of care ethics, which prioritize personal and concrete relationships with others (Zegura, DiSalvo, and Meng 2018). These participants made decisions by thinking about end users and the larger communities their work could harm or help. P9 described the care her team put into data handling as coming from a deeply personal sense of accountability: “We would have felt so bad if that data had been intercepted. We did not want to be the ones who caused that type of breach.”

P12 held a similar line when it came to speed: “Speed is a factor. But it cannot be done at the expense of good data. We can't choose speed over credibility.” P12 grew from working with a single native speaker to collaborating with a team of on-the-ground data collectors because he understood the scale required to responsibly represent low-resource languages:

I began to work with just one person, but again, I understood that this could be done at scale [...] if you want to build really good technology, you need to represent diversity. And if you want to build anything at scale, then you need to be able to collect a whole lot of data at approximately the same time.

Professional Community-Building Participants working within their institutions took time to build collective understanding over time (P1, P3, P4, P7, P8). P1 felt the potential to push against the data bias they cared about when they talked with their colleagues outside of work:

I have an inkling that my colleagues agree with me, but that's just because I know what their other values are based on other conversations [...] But in the professional environment, it's more of like, oh, we just talk about these as facts.

P15's data science and engineering teams developed a seamless, privacy-preserving, and responsible technical solution with roundtable discussions.

[Given the] concerns on what we should do with the IPs, somebody said, 'okay, if we use this process on hashing the IP, we'll get different values for each of the entries. We'll be able to count each of them independently, and will not have this compliance situation exposing the IPs.' He was a data engineer. He encrypted the IP, and once the table was ready, I performed some analysis on counting the different entries for the different timestamps and countries. I ran the model expecting to have the same marketing attribution rate.

As P2's workplace became increasingly hostile while being furloughed for six weeks and losing access to work communications, she found a crucial lifeline to communicate with colleagues outside of work.

I think everyone felt anxiety, and some grief, and maybe some depression; I don't know. One thing that was great, was just being able to keep in touch with

coworkers over things like Signal so that we have a space to communicate and support each other, and we don't have to worry about surveillance or something.

GenAI: An Emergent Moral Missions Case Study

The rapid corporatization of GenAI tools has introduced distinct market pressures that fundamentally complicate the moral missions our participants described (Lorrimar and Smartt 2026; Challenger, Gray, & Christmas Inc. 2026; Office of Management and Budget 2025). More than half of our participants raised GenAI as an emerging circumstance that evolved their moral missions. Unlike academic contexts, where communities of established and well-respected professionals openly dissent from unrestrained AI adoption (Bender and Hanna 2025), software engineers and technical practitioners in industry operate within incentive structures that actively reward, and in some cases, compel AI use (Wolfe and Mitra 2024; Connolly 2024; Tinsman, Krishan, and Neff 2026).

Defining GenAI Moral Missions

Even if participants didn't use GenAI for their responsible practice, most noted the unavoidable market push of GenAI across sectors. As P6 put it:

Whenever I get a new job, which has mostly been internships, they usually ask you to look at the organizational structure and think of what you can work on. [...] A lot of companies are not very sure of what they want to actually achieve with ML, and now we have LLMs. Many people are actually not sure. Well, a few people are sure. When they pull up RAG, at least you know they want to search a particular document or something. Sometimes there's a grey area, and they want experimentation.

While some participants embraced GenAI (P7, P11, P14), others explicitly questioned the hype. P4, a product manager collaborating with data scientists, shared her thoughts on the ongoing enthusiasm in her workplace to adopt GenAI.

My work opinion is [to be] cautious but experiment. My personal opinion is similar in that [...] be very, very cautious and experiment with the understanding that the data, the developmental process, to me, for these [GenAI technologies] were not ethical. No one had informed consent as to whether their stuff they put out on the internet was going into these; no one was paid for it, and no one even had the option to opt in or out.

Moral Missions Guiding GenAI Use

To P7's surprise, institutional interest in GenAI's potential for increased efficiency bolstered support for his effort to improve dataset documentation. Due to P7's measured GenAI approach, he not only exceeded institutional expectations but also met his responsibility standards.

When the new administration was coming in, we thought that it would change everything, but it was

quite the opposite. [...] They actually featured our dataset on their website and said that we are trying to make this process much faster and better with quality checks. It's not like we are making [it] faster, but we are dropping some things, or we are skipping some things; that's not the idea.

P13 did not personally use GenAI for his technique, but technical gaps in popular data visualization tools like Power BI led to his reluctant support of code generation. He instructs novice coders to: "go to Gemini, ask it to create a D3 page showing this data. Don't trust anything that it spits out, but check it, make sure that it's right."

P4 described staying engaged with workplace discussions of GenAI adoption, despite her own reservations, as a deliberate choice to stay in the decision-making room:

I still think that I have an obligation to test and do things and try to use them. Because the people that are staunchly against it and very vocal about it are not the ones invited into the room where decisions are being made for a business on whether or not to use them. But I'm also not going to be gung-ho, put it in everything, do everything like a lot of people.

With expertise in explainability, P11 explains how GenAI is technically misaligned with their approach to RDS practice.

Previously you would have a small set of data, then you would use pre-selected features, and then you would use it to predict something. So you had a lot of control over it. Whereas when we talk about LLMs based on transformer architecture, companies like OpenAI and Anthropic, they're literally using the whole internet. You just don't know where it's learning from [...] with LLMs, explainability, while critical, has become a really big challenge, because you just don't have a single source of information or corpus from which the information has been pegged.

Center Relationality to Nurture GenAI Moral Missions

Regardless of the decision to adopt or struggle against GenAI use, participants recognized their responsibility to maintain human oversight and engage their communities to scale responsible GenAI efforts. P7 built a pipeline that kept human judgment throughout the process:

We have developed a pipeline where we can ingest those PDFs. I first extract the text from that, and then use different LLMs, like Gemini 2.5 Flash, to externalize: who is the main federal agency responsible for this work? What is the location? We have developed some prompts to extract those and validate. [...] The initial annotation is human, then improving and scaling it up to maybe [millions of] pages is automatic. And then the final evaluation [is] again human. So, it's kind of like two places where humans are there.

P5's entire moral mission is equipping his local RDS community with the knowledge to advocate for African representation in GenAI. Therefore, he regularly engages with

the technology as a method for identifying African exclusion and was invigorated by the growing movement he plays a part in.

I'm talking about [GenAI in] my region; people are becoming more interested in [AI ethics]. It's not like before; people didn't really put much interest there. The awareness is taking momentum, and people are really coming into it and knowing that, okay, I have to speak, I have to advocate, I have to be included. So, it's making the whole thing easier now. I expect good inclusiveness in the future.

Across interviews, participants described GenAI not simply as a new technical tool, but as a site where organizational urgency, market pressure, professional identity, and moral conviction collided. Participants in our study did not all respond to GenAI the same way, with some refusing to use it altogether. Framing refusal as part of deeply held moral missions reveals that such practices are not simply *subtractive* stances against GenAI, but rather *generative* practices opening up new forms of relationality outside academic and capitalistic norms (Tuck and Yang 2014).

Moral Missions, the Augmented Undercommons, and Implications for Responsible Data Science Communities

The seemingly irreconcilable relationship between RDS practice and institutional logics is a persisting critique we surfaced in our study as well (Metcalf, Moss et al. 2019; Wong, Madaio, and Merrill 2023; Deng et al. 2022). We are the first to turn to the Black Studies theory of the *undercommons* (Moten and Harney 2013) to flag clearer directions for decoupling RDS practice from institutional logics. Moten and Harney (2013) articulate how professionals within academia subvert the often-exploitative conditions of their workplace to exercise their moral agency. While Moten and Harney (2013) first theorized the undercommons for the academy, Croskey, Benjamin, and Jacobs (2026) introduced the augmented undercommons as the first theoretical work to bring these ideas into HCI, describing how subversive designers operate within and against institutional constraints to advance justice-oriented work. We extend this theoretical lineage to illustrate how moral missions call for subversive data scientists, who **refuse credit, operate in parallel, and center collectivity** as the foundation of responsible practice (Croskey, Benjamin, and Jacobs 2026).

Refusing Credit: Define Moral Missions

The theme of *Domain Problematizing* reflects the idea of refusing credit as explored by Moten and Harney (2013) in the undercommons. The concept of credit represents the feeling that professionals ‘owe’ full cooperation to the institutions that employ them. In ‘paying off’ this credit, professionals tend to act in the best interests of their institutions (i.e., RDS appeals to reducing business risk (Wong, Madaio, and Merrill 2023; Metcalf, Moss et al. 2019)) even as these institutions uphold systems that harm and push communities to the margins. Participants struggled to exercise their

moral agency because they had institutional expectations in the back of their mind.

Moten and Harney (2013) contend that *refusing credit*, or letting go of the feeling of owing cooperation, can foster “fugitive publics” that “*operate outside the logic of institutional credit and repayment.*” Madaio et al. (2020) similarly identify “good tension” as a necessary catalyst for meaningful responsible AI discussions. When data scientists are able to view their professional responsibility as not in service to institutional ends but to societal progress, it may become easier to not only develop personal convictions about one’s professional domain but also to act on these convictions.

Operating in Parallel: Guide Moral Missions

Although working within institutional priorities can build buy-in for responsible techniques (Madaio et al. 2020), participants in our study recognize that this approval of their responsible practice comes at the cost of fully realizing their moral missions. They therefore *operated in parallel* by simultaneously translating community needs into professionalized spaces and challenging the boundaries of those same spaces (Croskey, Benjamin, and Jacobs 2026). Materially, participants risked losing clients (P9) and even discussed professional dissent outside of work (P1 and P2), establishing clear and often risky stances on the importance of centering communities within AI development regardless of the economic and social cost of such a stance.

Center Collectivity: Nurture Moral Missions

Our theme of *Professional Community-Building* also makes clear that refusal is hard to sustain alone. (Moten and Harney 2013) emphasize that building power requires study and planning in community, with colleagues, fellow professionals, and collaborators who share the stakes. When participants felt like the only ones concerned about moral dilemmas in their work, they embodied the principle of *collectivity* by forming connections with other subversive data scientists (Croskey, Benjamin, and Jacobs 2026). Collectivity is not only about sharing critiques but developing ad hoc strategies (Lin et al. 2026; So et al. 2025; Abdalla 2025). In line with the idea of the undercommons, sometimes these solutions can only be shared in privileged spaces and between colleagues with established relationships, but other times, the open sharing of RDS techniques is the catalyst for new fugitive publics.

Design Implications for RDS Tools

Our findings suggest that RDS tools should support not only the application of ethical principles, but also the moral decision-making procedures through which practitioners sustain those principles under institutional constraint. We therefore identify design implications for sociotechnical systems that help practitioners define, guide, and nurture moral commitments throughout data science work.

Develop moral agency Previous scholarship recommends embedding learning resources inside of RDS tools to support practitioners’ understanding of data ethics frameworks and theory so they can be appropriately applied to practice (Deng

et al. 2022; Wong, Madaio, and Merrill 2023). However, inspired by JafariNaimi, Nathan, and Hargraves (2015) and the rich moral analyses developed by our participants, we contend that even simple data ethics prompting can be powerful jumping-off points for ethical contextualization. The success of existing card-based tools supports this direction but often focuses on developing analyses on behalf of stakeholders (Friedman and Hendry 2012; Shen et al. 2021).

Instead, we imagine tools that build convictions and emotional investment for responsible practice may improve practitioners' capability to contextualize across projects (Dong et al. 2026; Madaio et al. 2020, 2024; Urovi et al. 2024). These tools can look like adapting existing value cards to probe personal moral analyses, appropriating short structured templates like positionality statements into journaling prompts for career-long introspection, or assessing practitioners' alignment to existing data ethics frameworks for meaningful investigation.

Maintain sociotechnical strategies Previous critiques of technical RDS toolkits highlight how generalizable techniques can be co-opted into ethics-washing and reinforce technosolutionist orientations that avoid the social dimensions of responsible practice (Selbst et al. 2019; Madaio et al. 2020; Wong, Madaio, and Merrill 2023). To combat the misappropriation of RDS techniques, previous scholarship called for embedding contextualized examples within toolkits and building avenues for non-technical stakeholders and community members to introduce socially-grounded context to technical considerations (Deng et al. 2022).

Our study reinforces critiques of purely technical RDS tools. Practitioners provided technical specifics in interviews, but the richest aspects of their technical narratives were the social considerations they made along the way; for example, determining stakeholder relationships to maintain privacy vision (P9), preparing communications to inform colleagues of responsible considerations (P1, P10, P11), and applying responsible scholarship to their practice (P7, P13). All of these demonstrate how technical and social action are deeply intertwined.

In addition to curating examples (Bako et al. 2023), we should develop tools to store their instances of responsible practice. Similar to repository-based platforms like GitHub, Kaggle, or Hugging Face, practitioners should be able to store sociotechnical artifacts of their responsible practices (Zhang, Muller, and Wang 2020). They can also keep track of what went well or what went wrong in each instance to retain and refine successful strategies to carry on to new contexts (Urovi et al. 2024). Explicit access control should be a major feature of these repositories, especially if these strategies are subversive. Practitioners should be able to decide to share with a few trusted allies, any professional who may find it useful, or keep it to themselves.

Build value-aligned communities Finally, the key to responsible practice revealed in our study and reinforced in existing scholarship is the role of community. Like any other task in data science work, RDS practices rely on a network of helpful colleagues for explicit guidance but encouragement as well (Pillai, Cha, and Wong 2026; Lin

et al. 2026). These networks take the form of public forums (e.g., Stack Overflow, Slack channels, in-person check-ins or workshops, accessible and secure channels like Signal) (Deng et al. 2022). Previous work calls on RDS tools to integrate professional online communities to support responsible practice (Deng et al. 2022). Sometimes, as highlighted by Widder (2024), there are no colleagues with similar values, and practitioners need to look outside of their institution to find like-minded professionals.

Therefore, we encourage tools to assess practitioners' values and support the identification of professionals with shared values (Widder et al. 2024). These relationships can set the stage for practitioners to swap notes, seek advice, collectively develop new strategies, and encourage sustained responsible practice through challenges (Wong, Madaio, and Merrill 2023). However, value-based communities can lead to echo chambers, or in the worst cases, the development of communities bound by harmful values (Lee et al. 2021). Therefore, moderation mechanisms should be in place to guard against the devolution of professional RDS communities.

Limitations

Qualitative research and phenomenological analysis, especially, should not be seen as generalizable or representative of our research inquiry. There is a vast range of RDS techniques and professional data science sectors that were not represented in our study. Additionally, connectivity issues impacted the flow of conversation and participants' ability to directly interact with the intended workspace/board. P8's full recording and transcript were lost, and therefore their analysis was constructed based on reflection board content, interview notes, facilitator notes, and confirmatory research. P3 and P10's reflection board interactions were not captured in video recordings, and therefore board interaction analyses were gathered from transcripts and interview notes.

Conclusion

We examine "*moral missions*" as phenomena of sustained responsible data science practice. In this paper, we present a semi-structured interview study with 15 responsible data scientists and AI practitioners to understand the moral decision-making procedures they use to articulate and actualize their convictions. We surface three interconnecting themes to capture our participants' approach to moral missions: 1) embodied introspection, 2) institutional circumvention, and 3) relationality. By unpacking the circumstances, moral analyses, and strategies of our participants, we also reveal pragmatic accounts of responsible GenAI decision-making. Finally, we contribute to the emerging scholarship of the (augmented) undercommons to call for subversive data science communities. In future work, we hope to co-design responsible data science tools with practitioners to realize virtual platforms and physical spaces for moral mission communities.

Ethics Statement

Given participants' professional commitments and obligations, we recognize the potential risks associated with discussing personal values, workplace practices, and organizational environments. Accordingly, participant privacy and maintaining confidentiality were prioritized throughout the research process with measures such as anonymization and data de-identification. Participants were offered opportunities to review recordings and transcripts, request the anonymization or removal of specific information, and withdraw from the study at any time with their data excluded. We also recognize that practices and strategies described in this work may have implications beyond the specific contexts within which they are described. While our focus is on those who identify as responsible data scientists and seek to enact values in service of social good, under different circumstances, approaches described in this work can be employed in ways that are exclusionary, harmful, or contrary to social good and public interest. Accordingly, we place significant attention on discussing implications of our findings and the directions they suggest for supporting RDS and RDS communities.

Positionality Statement

The lead researcher is a Computer Science PhD student with experience in responsible data science as an intern, research assistant, and student. While this experience provided familiarity with responsible data science concepts and practices, she remains removed from the everyday experiences of many practicing data scientists. Given her research focus in data ethics, she remained attentive to the possibility of imposing theoretical frameworks onto participants' accounts and therefore sought to ground interview questions and follow-up discussions in participants' lived experiences. The lead researcher, along with several members of the research team, identifies with groups underrepresented in data science and computing. This informed a commitment to recruiting participants with diverse backgrounds and ensuring space for perspectives that may be less visible within data science research and practice.

The broader research team brings expertise, interests, and experience spanning computer science and information science, including responsible data science, data and AI ethics, human-computer interaction, data management and visualization, and AI policy. Team members bring varying levels of direct experience with data science practice, ranging from former non-professional practitioners and industry collaborators to researchers working in adjacent areas. We recognize that our disciplinary backgrounds, experiences, and commitments shape both our approach to research and the interpretations we develop. Throughout the research process, we sought to remain attentive to participants' perspectives and create space for them to describe their experiences in their own terms.

Acknowledgments

This work was supported in part by the National Science Foundation (Awards # 2402718, # 2514565, # 2141506).

References

- Abdalla, M. 2025. \$100,000 or the Robot Gets It! Tech Workers' Resistance Guide: Tech Worker Actions, History, Risks, Impacts, and the Case for a Radical Flank. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 8(1): 2–14.
- Ajmani, L.; Stapleton, L.; Houtti, M.; and Chancellor, S. 2024. Data Agency Theory: A Precise Theory of Justice for AI Applications. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '24, 631–641. New York, NY, USA: Association for Computing Machinery. ISBN 9798400704505.
- Ali, S. J.; Christin, A.; Smart, A.; and Katila, R. 2023. Walking the walk of AI ethics: Organizational challenges and the individualization of risk among ethics entrepreneurs. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, 217–226.
- Alvarado Garcia, A.; Wong-Villacres, M.; Hernández, B.; and Le Dantec, C. A. 2024. Bitacora: A Toolkit for Supporting NonProfits to Critically Reflect on Social Media Data Use. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24. New York, NY, USA: Association for Computing Machinery. ISBN 9798400703300.
- Bako, H. K.; Liu, X.; Battle, L.; and Liu, Z. 2023. Understanding how Designers Find and Use Data Visualization Examples. *IEEE Transactions on Visualization and Computer Graphics*, 29(1): 1048–1058.
- Barrett, T.; Okolo, C. T.; Biira, B.; Sherif, E.; Zhang, A.; and Battle, L. 2025. African Data Ethics: A Discursive Framework for Black Decolonial AI. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '25, 334–349. New York, NY, USA: Association for Computing Machinery. ISBN 9798400714825.
- Bellamy, R. K. E.; Dey, K.; Hind, M.; Hoffman, S. C.; Houde, S.; Kannan, K.; Lohia, P.; Martino, J.; Mehta, S.; Mojsilovic, A.; Nagar, S.; Ramamurthy, K. N.; Richards, J.; Saha, D.; Sattigeri, P.; Singh, M.; Varshney, K. R.; and Zhang, Y. 2018. AI Fairness 360: An Extensible Toolkit for Detecting, Understanding, and Mitigating Unwanted Algorithmic Bias. ArXiv:1810.01943 [cs].
- Bender, E.; and Hanna, A. 2025. *The AI Con: How to Fight Big Tech's Hype and Create the Future We Want*. Harper-Collins. ISBN 9780063418554.
- Bennett, S. 2025. Artificial Intelligence and the ethics of navigating ambiguity. *Big Data & Society*, 12(2): 20539517251347594.
- Berne, P.; Morales, A. L.; Langstaff, D.; and Sins Invalid. 2018. Ten Principles of Disability Justice. *WSQ: Women's Studies Quarterly*, 46(1): 227–230.
- Bhattacharjee, S.; and Ribes, D. 2025. Technical Responses To Critique: The Case Of Skin Tone. 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.

- Bird, S.; Dudík, M.; Edgar, R.; Horn, B.; Lutz, R.; Milan, V.; Sameki, M.; Wallach, H.; Walker, K.; and Design, A. 2022. Fairlearn: A toolkit for assessing and improving fairness in AI.
- Birhane, A. 2021. Algorithmic injustice: a relational ethics approach. *Patterns*, 2(2): 100205.
- Boenig-Liptsin, M.; Tanweer, A.; and Edmundson, A. 2022. Data Science Ethos Lifecycle: Interplay of Ethical Thinking and Data Science Practice. *Journal of Statistics and Data Science Education*, 30(3): 228–240.
- Borning, A.; and Muller, M. 2012. Next steps for value sensitive design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, 1125–1134. New York, NY, USA: Association for Computing Machinery. ISBN 9781450310154.
- Challenger, Gray, & Christmas Inc. 2026. Challenger Report: December 2025. Technical report, Challenger, Gray, & Christmas, Inc. <https://www.challengergray.com/wp-content/uploads/2026/01/Challenger-Report-December-2025.pdf>.
- Connolly, R. 2024. Public Computing Intellectuals in the Age of AI Crisis. *ACM Trans. Comput. Educ.*, 24(4).
- Croskey, P.; Benjamin, R.; and Jacobs, J. 2026. The Augmented Undercommons: A Framework for Liberatory HCI Research within Ethically Compromised Institutions. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*, CHI '26. New York, NY, USA: Association for Computing Machinery. ISBN 9798400722783.
- Dahlberg, H.; and Dahlberg, K. 2020. Open and Reflective Lifeworld Research: A Third Way. *Qualitative Inquiry*, 26(5): 458–464.
- Deng, W. H.; Nagireddy, M.; Lee, M. S. A.; Singh, J.; Wu, Z. S.; Holstein, K.; and Zhu, H. 2022. Exploring How Machine Learning Practitioners (Try To) Use Fairness Toolkits. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '22, 473–484. New York, NY, USA: Association for Computing Machinery. ISBN 9781450393522.
- D'Ignazio, C.; and Klein, L. F. 2020. *Data Feminism*. MIT Press.
- Dong, Z.; Patil, A.; Shoda, Y.; Battle, L.; and Wall, E. 2025. Behavior Matters: An Alternative Perspective on Promoting Responsible Data Science. *Proc. ACM Hum.-Comput. Interact.*, 9(2).
- Dong, Z.; Wu, K.; Battle, L.; and Wall, E. 2026. Evaluating Behavior Change Interventions for Responsible Data Science. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*, 1–15.
- Eatough, V.; and Smith, J. A. 2017. Interpretative phenomenological analysis. *The Sage handbook of qualitative research in psychology*, 2: 193–209.
- Floridi, L.; Cows, J.; Beltrametti, M.; Chatila, R.; Chazerand, P.; Dignum, V.; Luetge, C.; Madelin, R.; Pagallo, U.; Rossi, F.; et al. 2018. AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and machines*, 28(4): 689–707.
- Fortin, C.; Hennessy, K.; Baur, R.; and Fortin, P. 2013. Beyond the vision paradigm: design strategies for crossmodal interaction with dynamic digital displays. In *Proceedings of the 2nd ACM International Symposium on Pervasive Displays*, PerDis '13, 91–96. New York, NY, USA: Association for Computing Machinery. ISBN 9781450320962.
- Friedman, B.; and Hendry, D. 2012. The envisioning cards: a toolkit for catalyzing humanistic and technical imaginations. In *Proceedings of the SIGCHI conference on human factors in computing systems*, 1145–1148.
- Friedman, B.; Kahn, P.; Borning, A.; Zhang, P.; and Galletta, D. 2006. *Value Sensitive Design and Information Systems*. M. E. Sharpe. ISBN 978-94-007-7843-6.
- Friedman, B.; and Kahn, P. H. 2002. *Human values, ethics, and design*, 1177–1201. USA: L. Erlbaum Associates Inc. ISBN 0805838384.
- Geburu, T.; Morgenstern, J.; Vecchione, B.; Vaughan, J. W.; Wallach, H.; Iii, H. D.; and Crawford, K. 2021. Datasheets for datasets. *Communications of the ACM*, 64(12): 86–92.
- Heger, A. K.; Passi, S.; Dhanorkar, S.; Kahn, Z.; Wang, R.; and Vorvoreanu, M. 2025. Towards a Responsible AI Organizational Maturity Model. *Proc. ACM Hum.-Comput. Interact.*, 9(7).
- Hertweck, C.; Heitz, C.; and Loi, M. 2024. What's Distributive Justice Got to Do with It? Rethinking Algorithmic Fairness from a Perspective of Approximate Justice. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 7(1): 597–608.
- Horvath, A.-S.; and Raptis, D. 2025. How Students in Creative Educations Appropriate Technology: A phenomenological analysis. 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.
- Hsu, J.; Ravichandran, R.; Zhang, E.; and Keung, C. 2021. Open Data Standard and Analysis Framework: Towards Response Equity in Local Governments. In *Equity and Access in Algorithms, Mechanisms, and Optimization*, 1–8. – NY USA: ACM. ISBN 978-1-4503-8553-4.
- JafariNaimi, N.; Nathan, L.; and Hargraves, I. 2015. Values as Hypotheses: Design, Inquiry, and the Service of Values. *Design Issues*, 31(4): 91–104.
- Kallia, M. 2025. To Be, or to Be Otherwise? Silicon Souls in Search of Dasein and Authentic Human Engagement in the Age of Generative AI. In *Proceedings of the 2025 ACM Conference on International Computing Education Research V.1*, ICER '25, 240–255. New York, NY, USA: Association for Computing Machinery. ISBN 9798400713408.
- Klein, L.; and D'Ignazio, C. 2024. Data Feminism for AI. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '24, 100–112. New York, NY, USA: Association for Computing Machinery. ISBN 9798400704505.
- Krijger, J.; Thuis, T.; De Ruiter, M.; Ligthart, E.; and Broekman, I. 2023. The AI ethics maturity model: a holistic approach to advancing ethical data science in organizations. *AI and Ethics*, 3(2): 355–367.

- Kuehnert, B.; Kim, R.; Forlizzi, J.; and Heidari, H. 2025. The “Who”, “What”, and “How” of Responsible AI Governance: A Systematic Review and Meta-Analysis of (Actor, Stage)-Specific Tools. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*, FAccT ’25, 2991–3005. New York, NY, USA: Association for Computing Machinery. ISBN 9798400714825.
- Lee, C.; Yang, T.; Inchoco, G. D.; Jones, G. M.; and Satyanarayan, A. 2021. Viral Visualizations: How Coronavirus Skeptics Use Orthodox Data Practices to Promote Unorthodox Science Online. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI ’21. New York, NY, USA: Association for Computing Machinery. ISBN 9781450380966.
- Lin, H.; Lisnic, M.; Akbaba, D.; Meyer, M.; and Lex, A. 2026. Here’s what you need to know about my data: Exploring Expert Knowledge’s Role in Data Analysis. *IEEE Transactions on Visualization and Computer Graphics (VIS)*, 32.
- Lin, W.; Wang, J.-Y.; and Yueh, H.-P. 2022. Learning Information Ethical Decision Making With a Simulation Game. *Frontiers in Psychology*, Volume 13 - 2022.
- Linder, J.; and Arvola, M. 2017. IPA in UX Research: Interpretative Phenomenological Analysis in a User Experience Design Practice. In *Proceedings of the European Conference on Cognitive Ergonomics*, ECCE ’17, 17–24. New York, NY, USA: Association for Computing Machinery. ISBN 9781450352567.
- Lipton, Z. C. 2017. The Mythos of Model Interpretability. ArXiv:1606.03490 [cs].
- Lorrimar, V.; and Smartt, T. 2026. Silicon Valley’s AI ‘tokenmaxxing’ obsession has a big problem—and philosophers saw it coming. *The Conversation*.
- Lundberg, S.; and Lee, S.-I. 2017. A Unified Approach to Interpreting Model Predictions. ArXiv:1705.07874 [cs].
- Madaio, M.; Egede, L.; Subramonyam, H.; Wortman Vaughan, J.; and Wallach, H. 2022. Assessing the fairness of AI systems: AI practitioners’ processes, challenges, and needs for support. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW1): 1–26.
- Madaio, M. A.; Chen, J.; Wallach, H.; and Wortman Vaughan, J. 2024. Tinker, Tailor, Configure, Customize: The Articulation Work of Contextualizing an AI Fairness Checklist. *Proc. ACM Hum.-Comput. Interact.*, 8(CSCW1).
- Madaio, M. A.; Stark, L.; Wortman Vaughan, J.; and Wallach, H. 2020. Co-Designing Checklists to Understand Organizational Challenges and Opportunities around Fairness in AI. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI ’20, 1–14. New York, NY, USA: Association for Computing Machinery. ISBN 9781450367080.
- 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.
- Midha, S.; Wilson, M. L.; and Sharples, S. 2022. Lived Experiences of Mental Workload in Everyday Life. 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.
- Mitchell, M.; Wu, S.; Zaldivar, A.; Barnes, P.; Vasserman, L.; Hutchinson, B.; Spitzer, E.; Raji, I. D.; and Gebru, T. 2019. Model cards for model reporting. In *Proceedings of the conference on fairness, accountability, and transparency*, 220–229.
- Mohamed, S.; Png, M.-T.; and Isaac, W. 2020. Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence. *Philosophy & Technology*, 33(4): 659–684.
- Morley, J.; Elhalal, A.; Garcia, F.; Kinsey, L.; Mökander, J.; and Floridi, L. 2021. Ethics as a Service: A Pragmatic Operationalisation of AI Ethics. *Minds and Machines*, 31(2): 239–256.
- Moten, F.; and Harney, S. 2013. The undercommons: Fugitive planning and black study.
- Murray-Rust, D.; Lupetti, M. L.; Nicenboim, I.; and Hoog, W. v. d. 2023. Grasping AI: experiential exercises for designers. *AI Soc.*, 39(6): 2891–2911.
- Naeem, M.; Ozuem, W.; Howell, K.; and Ranfagni, S. 2023. A step-by-step process of thematic analysis to develop a conceptual model in qualitative research. *International journal of qualitative methods*, 22: 16094069231205789.
- Office of Management and Budget. 2025. 2025 Federal Agency Artificial Intelligence (AI) Use Case Inventory. <https://github.com/ombegov/2025-Federal-Agency-AI-Use-Case-Inventory/tree/main>.
- Pillai, A. G.; Cha, I.; and Wong, R. Y. 2026. Navigating Organizational Infrastructures: Addressing Ethics Uncertainty Experienced by Middle Managers in AI Development. In *Proceedings of the 2026 ACM Conference on Fairness, Accountability, and Transparency*, 570–590. Montreal QC Canada: ACM. ISBN 979-8-4007-2596-8.
- Preti, A. A. 2018. Developing Habits of Moral Reflection: Dewey, Moral Inquiry, and Practical Ethics. In Englehardt, E. E.; and Pritchard, M. S., eds., *Ethics Across the Curriculum—Pedagogical Perspectives*, 147–163. Cham: Springer International Publishing. ISBN 978-3-319-78939-2.
- Rizvi, N.; Smith, T.; Vidyala, T.; Bolds, M.; Strickland, H.; Begel, A.; Williams, R.; and Munyaka, I. 2025. “I Hadn’t Thought About That”: Creators of Human-like AI Weigh in on Ethics & Neurodivergence. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*, FAccT ’25, 3385–3399. New York, NY, USA: Association for Computing Machinery. ISBN 9798400714825.
- Saltz, J. S.; and Dewar, N. 2019. Data science ethical considerations: a systematic literature review and proposed project framework. *Ethics and Information Technology*, 21(3): 197–208.
- Schwartz, S. H. 1994. Are There Universal Aspects in the Structure and Contents of Human Values? *Journal of Social Issues*, 50(4): 19–45.

- 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. Atlanta GA USA: ACM. ISBN 978-1-4503-6125-5.
- Shen, H.; Deng, W. H.; Chattopadhyay, A.; Wu, Z. S.; Wang, X.; and Zhu, H. 2021. Value Cards: An Educational Toolkit for Teaching Social Impacts of Machine Learning through Deliberation. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '21, 850–861. New York, NY, USA: Association for Computing Machinery. ISBN 9781450383097.
- Shen, H.; Kneare, T.; Ghosh, R.; Yang, Y.-J.; Clark, N.; Mitra, T.; and Huang, Y. 2025. ValueCompass: A Framework for Measuring Contextual Value Alignment Between Human and LLMs. In Zhang, C.; Allaway, E.; Shen, H.; Miculicich, L.; Li, Y.; M'hamdi, M.; Limkonchotiawat, P.; Bai, R. H.; T.y.s.s., S.; Han, S. S.; Thapa, S.; and Rim, W. B., eds., *Proceedings of the 9th Widening NLP Workshop*, 75–86. Suzhou, China: Association for Computational Linguistics. ISBN 979-8-89176-351-7.
- Shipp, S.; LaLonde, D.; and Martinez, W. 2023. Making Ethical Decisions Is Hard! *CHANCE*, 36(4): 42–50.
- Smith, H. M. 2018. Using Moral Principles to Guide Decisions. In *Making Morality Work*. Oxford University Press. ISBN 9780199560080.
- So, S.; Sou, V.; Munson, S. A.; and Ghoshal, S. 2025. The Cruel Optimism of Tech Work: Tech Workers' Affective Attachments in the Aftermath of 2022-23 Tech Layoffs. 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.
- Stoudt, S.; Jernite, Y.; Marshall, B.; Marwick, B.; Sharan, M.; Whitaker, K.; and Danchev, V. 2024. Ten simple rules for building and maintaining a responsible data science workflow. *PLOS Computational Biology*, 20(7): e1012232.
- Tinsman, C.; Krishan, A.; and Neff, G. 2026. Navigating AI at Work: An Ethics-in-Practice Approach. In *Proceedings of the 2026 ACM Conference on Fairness, Accountability, and Transparency*, 5726–5755. Montreal QC Canada: ACM. ISBN 979-8-4007-2596-8.
- Tuck, E.; and Yang, K. W. 2014. R-Words: Refusing Research. In Paris, D.; and Winn, M. T., eds., *Humanizing Research: Decolonizing Qualitative Inquiry with Youth and Communities*, chapter 12, 223–248. City Road, London: SAGE Publications.
- Urovi, V.; Celebi, R.; Sun, C.; Kumar, P.; Rieswijk, L.; Erard, M.; Yilmaz, A.; Moodley, K.; and Dumontier, M. 2024. TAPS Responsibility matrix: a tool for responsible data science by design. *Journal of Responsible Innovation*, 11(1): 1–32.
- Varanasi, R. A.; and Goyal, N. 2023. “It is currently hodgepodge”: Examining AI/ML practitioners' challenges during co-production of responsible AI values. In *Proceedings of the 2023 CHI conference on human factors in computing systems*, 1–17.
- Wang, A.; Datta, T.; and Dickerson, J. P. 2024. Strategies for Increasing Corporate Responsible AI Prioritization. AAAI. In *ACM Conference on AI, Ethics, and Society (AIES)*.
- 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.
- Widder, D. G.; Dabbish, L.; Herbsleb, J. D.; and Martelaro, N. 2024. Power and Play: Investigating “License to Critique” in Teams' AI Ethics Discussions. *Proc. ACM Hum.-Comput. Interact.*, 8(CSCW2).
- Winner, L. 1980. Do Artifacts Have Politics? *Daedalus*, 109(1): 121–136.
- Wolfe, R.; and Mitra, T. 2024. The Implications of Open Generative Models in Human-Centered Data Science Work: A Case Study with Fact-Checking Organizations. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 7(1): 1595–1607.
- Wong, R. Y.; Madaio, M. A.; and Merrill, N. 2023. Seeing Like a Toolkit: How Toolkits Envision the Work of AI Ethics. *Proc. ACM Hum.-Comput. Interact.*, 7(CSCW1).
- Yoo, D.; Logler, N.; Ballard, S.; and Friedman, B. 2022. Multi-lifespan Envisioning Cards: Journeying from Design Theory to Tools for Action. In *Proceedings of the 2022 ACM Designing Interactive Systems Conference*, DIS '22, 557–570. New York, NY, USA: Association for Computing Machinery. ISBN 9781450393584.
- Zegura, E.; DiSalvo, C.; and Meng, A. 2018. Care and the Practice of Data Science for Social Good. In *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*, COMPASS '18. New York, NY, USA: Association for Computing Machinery. ISBN 9781450358163.
- Zhang, A. X.; Muller, M.; and Wang, D. 2020. How do Data Science Workers Collaborate? Roles, Workflows, and Tools. *Proc. ACM Hum.-Comput. Interact.*, 4(CSCW1).