RAI Advocacy: Communicative Strategies for Advancing Responsible AI in Large Technology Companies

Jordan Duran[1], Samir Passi[2], Mihaela Vorvoreanu[2]
[1]UNC Charlotte
[2]Microsoft
jduran4@charlotte.edu, v-sapassi@microsoft.com, mihaela.vorvoreanu@microsoft.com

Abstract

Despite perceived tensions between Responsible AI (RAI) and business objectives in large technology companies, RAI efforts still advance thanks to the persistent, often invisible work of passionate advocates who take on RAI work, often in addition to their formal roles. In this paper, we examine the work of such RAI advocates from an organizational communication perspective that enables us to understand how organizational realities are continuously constructed and negotiated through communication. Specifically, we look at RAI advocates' communicative moves – the advocacy strategies they use to address RAI challenges and get RAI work done. Through an analysis of 22 in-depth interviews with RAI advocates, we identify shared obstacles related to getting buy-in for and facilitating RAI work, and 14 distinct communicative strategies advocates use to address them. Our findings highlight the demanding yet under-recognized labor that enables and shapes RAI work within organizations. We conclude by discussing how organizations can better support RAI advocacy efforts and, ultimately, envision a future where RAI is normalized in everyday organizational processes such that advocacy is no longer needed.

1 Introduction

Despite the growing consensus on the importance of responsible artificial intelligence (RAI) and regulatory pressures, large technology companies often face significant challenges in implementing RAI principles (Mittelstadt 2019; Schiff et al. 2020). Recent scholarship has highlighted many challenges to accomplishing RAI work in organizations, ranging from a lack of leadership support and resourcing to difficulties working in an ambiguous space with imperfect solutions (Ali et al. 2023; Widder et al. 2023; Rakova et al. 2021; Madaio et al. 2024; Schiff et al. 2020). However, even when not prioritized, RAI work does happen in technology companies, albeit to varying degrees of success. Even in the face of pushback, passionate advocates work hard to champion RAI and ensure some level of RAI work happens within their organizations.

In this paper, we adopt an organizational communication theoretical lens to understand the work of RAI advocacy. This theoretical lens views organizations not as static monoliths, but as continually *organized* through communication: being continuously (re)constituted through the communicative acts of people, some of whom have more or less power than others. Without people communicating – ideating, negotiating, (dis)agreeing – organizations, their products, services, values, practices, and cultures would not exist. While it is convenient to treat corporations as monoliths, in reality, decisions are made by people and reflect the conversations and tensions they encounter in everyday work. What conversations and tensions do people have in large technology companies about RAI? What does the work of advocating for RAI look like? What communicative strategies do RAI advocates use? What challenges do they encounter? Such questions lie at the heart of this paper.

We address these questions through an in-depth exploration of RAI advocacy at a large technology company, followed by triangulation with a few participants from other similar companies. We conducted in-depth semi-structured interviews with 22 self-identified RAI advocates, sixteen of which worked at the same technology company. In doing so, we illuminate the invisible work (Hatton 2017; Star and Strauss 1999) of RAI advocacy, including, but not limited to, raising awareness for RAI, getting buy-in for RAI work, and helping teams address RAI risks in products while staying humble and diffusing adversarial attitudes. We identify fourteen communicative strategies RAI advocates use to address five major challenges and shed light on the emotional and mental toll of doing this often unpaid and unrecognized work. We discuss implications for future research and for supporting RAI advocacy in organizations committed to advancing RAI.

Our goal is to take a necessary first step in advancing our understanding of the under-studied work of RAI advocacy through an in-depth exploration in the context of one large technology company. In technology companies, RAI advocacy often happens in response to push-back at the interpersonal, team, and organizational levels. Conversations on RAI advocacy often involve discussing sensitive topics such as "unreasonable" colleagues, "bad" organizations, or leadership that "doesn't care." Corporate employees hesitate to discuss such matters, particularly with people from other companies. Two of the authors work at a large technology company and have good rapport with its RAI practitioners, sixteen of whom trusted us with candid conversations. We conducted six additional interviews with self-identified RAI advocates at other companies, whom we identified through snowball sampling. In these additional interviews, we were interested to see whether RAI advocacy differs across organizational contexts (data triangulation). We were surprised to notice very similar themes, so we report findings across the two groups together.

With this research, we aim for transferability of findings. Other employees at various companies who engage in RAI advocacy and experience similar challenges as our participants might benefit from learning about the communicative strategies our interviewees shared. As with any in-depth qualitative research, we do not aim for nor claim that these findings describe RAI advocacy across the technology industry. Instead, we share "lessons learned" that others in similar situations may benefit from.

2 Background and Related Work

Our approach to RAI advocacy has its roots in critical scholarship on AI and RAI, especially work on the sociotechnical and organizational aspects of RAI. We build on this understanding of RAI advocacy using the analytical lens of *communication as constitutive of organization* (CCO) from the field of organizational communication.

2.1 RAI Advocacy

Despite the growing consensus around the need for RAI and the increasing number of RAI tools, organizations struggle to implement RAI in practice. Multiple barriers hamper RAI work, such as a lack of buy-in (Ali et al. 2023; Rakova et al. 2021), organizational silos (Piorkowski et al. 2021; Rakova et al. 2021; Jing et al. 2024), a lack of incentives (Widder et al. 2023; Deng et al. 2023), communication challenges (Deng et al. 2022; Piorkowski et al. 2021; Nahar et al. 2022), and mismatched expectations and values among stakeholders (Madaio et al. 2024; Deng et al. 2023; Holstein et al. 2019; Jakesch et al. 2022; Piorkowski et al. 2021). The interdisciplinary nature of RAI work further requires "meaningful collaboration" (Madaio et al. 2024) across many disciplines with distinct expertise, goals, and vocabularies (Madaio et al. 2024; Schiff et al. 2020; Deng et al. 2023; Piorkowski et al. 2021; Nahar et al. 2022). Such collaboration is challenging in organizations, especially given the uneven power dynamics between disciplines (Ali et al. 2023). The fact that organizations prioritize efficiency and profitability exacerbates these challenges, creating substantial resistance to RAI work that is often perceived as complicating or even stifling progress (Whittlestone et al. 2019). The ideals of RAI often appear misaligned with the everyday realities of organizations (Ali et al. 2023).

Despite these substantial barriers, RAI work does happen in organizations (Madaio et al. 2024; Ali et al. 2023; Widder et al. 2023; Schiff et al. 2020; Deng et al. 2023; Madaio et al. 2020). This is primarily made possible by the critical work of **RAI advocacy**—the concerted and persuasive effort to advance RAI in organizations by strategically addressing RAI's sociotechnical and organizational challenges. Recent research shows that for RAI to happen, "passionate individual advocates" (Madaio et al. 2020) must regularly act as "ethics entrepreneurs" (Ali et al. 2023) to enable value alignment and reduce frictions in the "understanding, contextualizing, and evaluation" (Deng et al. 2023) of RAI problems and solutions in organizations. RAI advocates take on this crucial work, often outside their job description, to position RAI not as a roadblock but as an enabler of successful AI products (Nedzhvetskaya and Tan 2024). This requires them to, among other things, raise awareness, build coalitions, and navigate institutional inertia and power dynamics (Widder et al. 2023; Madaio et al. 2024; Schiff et al. 2020; Deng et al. 2023; Madaio et al. 2020). While RAI advocacy is crucial to the success of RAI in organizations, we know little about what constitutes on-the-ground RAI advocacy work, how it is done, what strategies it entails. In this paper, we take a first step to bridge this gap, focusing on the communicative work of RAI advocates.

2.2 Seeing RAI Through an Organizational Communication Lens

To study the everyday work of RAI advocates within organizations, we draw on research in organizational communication, which provides a theoretical background to explore how tensions and challenges shape organizations through ongoing communicative activity. In this view, organizations are replaced with *organizing* - continually constructing organizations through their members' communicative acts (Weick, Sutcliffe, and Obstfeld 2005; Weick 1979). Rather than viewing communication as discrete messages separate from the issues they describe, the view of organizations being constituted through communication argues that messages and the issues they describe are one and the same, to the extent that they shape how organizational members develop understandings, values, and courses of action (Ashcraft, Kuhn, and Cooren 2009; Schoeneborn, Kuhn, and Kärreman 2019; Brummans et al. 2024). RAI activities – deciding what constitutes a risk, what risk level is acceptable, brainstorming and negotiating mitigations, coordinating their implementations – and, at a fundamental level, persuading people that RAI matters – are all communicative acts that continuously constitute *organizing* around RAI. It is through communication that members collectively negotiate what RAI is, how it is done, how much it is valued, and where it stands among an organization's priorities.

In fact, the vast majority of work, including building technology, is communication. Prior scholarship has articulated how organizational communication facilitates the social construction of technology, such as through technological framing and sociotechnical problem construction (Leonardi 2011, 2015; Mazmanian, Cohn, and Dourish 2014; Orlikowski and Scott 2015; Orlikowski and Gash 1994). Furthermore, as Bisel et al. (Bisel et al. 2011) articulate, organizational members collectively negotiate understandings of ethics through communication, as "communication itself is the behavior that imbues workplace (ethical or unethical) behavior with value" (p. 154).

The dynamics of negotiating and constituting organizational realities through communication are not always smooth. Tensions and conflicts occur, but they can be transformative when approached constructively. For instance, the visionary early 20th century theorist Mary Parker Follett saw conflict as a natural friction that occurred with the ongoing introduction of new ideas and cooperation within organizations. In particular, Follett wrote of resolving differences through integration, which she identified as finding resolution where opposing parties could creatively work together toward an alternative solution without compromise or resort to coercive tactics. She argued that this type of cooperative resolution via integration could promote "power-with" (Follett 1925; Berman and Van Buren 2015), where decision-making could be approached collectively within organizations rather than stemming from the authority held in particular roles (Mumby and Kuhn 2024).

Despite the opportunity that organizational tensions may pose for cooperation and change, they are still challenging to navigate in everyday situations. Putnam et al. (Putnam, Fairhurst, and Banghart 2016) define tensions in organizations as times when persistent topics or situations where there is stress, anxiety, discomfort, or tightness in making choices are enacted in communication. Tensions are a result of navigating multiple, sometimes contradictory discourses that RAI advocates frequently face when balancing various business demands and RAI goals (Widder 2024; Browne, Drage, and McInerney 2024).

Persistent tensions are powerful within organizations because they can disincentivize speaking up, especially around moral or ethical concerns. The reluctance to openly discuss ethics issues or raise moralized objections at work has been characterized as the "moral-mum effect," which is the result of both socialization processes that guide employees to not voice moral or ethical concerns, and impression management norms that make moral or ethical concerns difficult to discuss in everyday interactions (Bisel et al. 2011; Bisel 2018). In other words, it is difficult to "call out" others as being unethical, especially if one has a lower status, which is frequently the case for practitioners who focus on ethics in AI work (Widder 2024). Moral or ethical dissent in these contexts thus has a face-threatening nature which serves to further silence upward ethical dissent or sometimes creates situations where issues are labeled as benign rather than unethical (Kassing 2002; Zanin, Bisel, and Adame 2016).

However, despite the challenges related to discussing ethical issues in everyday organizational talk, regularly raising ethical issues encourages others to broach such issues, lowering the face-threatening nature of moral or ethical talk. Zanin et al. (Zanin, Bisel, and Adame 2016) conceptualize the spread of ethical talk as "moral contagion" and find that moral talk from leadership further facilitates ethical dissent and trust among employees. More broadly, organizations that craft supportive cultures for dissent or upward advocacy from employees can realize benefits at both the individual and organizational level. Support for employee dissent, advocacy, or issue-selling has been linked to improvements in organizational decision-making (Garner 2013; Garner and Ragland 2019; Garner and Garner 2011; Hegstrom 1990; Dutton et al. 2001), creativity (Kajzer Mitchell and Walinga 2017; Lu et al. 2019), fostering successful change initiatives (Detert and Treviño 2010; Satterstrom, Kerrissey, and DiBenigno 2021), and promoting social equity (Piderit and Ashford 2003; Kundro, Neely, and Muir 2024).

Given this background, a key question is how can individuals who care about RAI communicate concerns in ways that change existing ideas or offer new perspectives, despite tensions. This line of thinking follows a dialectical analysis of resistance within organizations that focuses on *understanding how individuals navigate and influence ongoing power dynamics and practices* (Mumby et al. 2017; Mumby 2005). In this work, we draw upon these theoretical perspectives with the aim of understanding how individuals who advocate for RAI navigate and influence ongoing power dynamics and practices related to getting RAI work done. What communicative strategies do RAI advocates employ? What challenges do they encounter? What does the communicative work of getting RAI done consist of, what are its rewards and costs?

3 Research Methods

To address these research questions, we conducted an in-depth interview study with 22 employees of large technology organizations who self-identified as RAI advocates. We began by interviewing sixteen participants who were employed at the same company as two of the authors, aiming for in-depth, candid conversations. These participants were familiar with our research group's reputation and, although they asked for reassurance regarding confidentiality, engaged in discussion of difficult topics. To achieve data triangulation, we then interviewed six participants from other organizations, whom we identified through snowball sampling.

3.1 Participants

We used criterion sampling to select participants who worked in the information technology field, preferably at large multinational technology companies, and who self-identified as RAI advocates. We advertised the research on discussion groups and email lists, and through snowball sampling, first inside one company, then others. We included participants who had RAI in their formal job descriptions, participants who did volunteer RAI work, and participants who advocated for RAI even though it was not a part of their role.

Twenty-two RAI advocates participated in our interviews. Many participants expressed concerns about confidentiality and for this reason, we report only the information about them needed to establish context for the findings. All but two participants were employed by large multinational technology companies. Two were employed at technology-related non-profit organizations. Ten participants worked in the product management discipline, in various roles related to product, research, or customer success. The remaining 12 participants worked in UX (UX design, UX research; n=4), research (as applied or research scientists; n=4), engineering/data science (n=3), and other (n=1). Of the 22 participants, eight had RAI in their official job descriptions. They worked in research teams that focused on RAI, or were part of their organization's RAI programs. Six participants did not have RAI in their official job descriptions, but did formal

RAI activities such as training and advising product teams on a volunteer basis. The remaining eight participants advocated for RAI in their work as RAI conversations or issues came up. All participants were based in the United States or Canada.

As described earlier, sixteen of our 22 participants were employed at the same technology company. We first interviewed these sixteen participants from the same company, and left those from other companies towards the end, so we could better assess data saturation. We reached data saturation about halfway through the interviews, but we continued interviewing RAI advocates for added confidence in our findings. Indeed, participants from the other organizations expressed eerily similar concerns, giving us confidence data saturation had been reached. Therefore, we do not separate findings per group.

3.2 Data Collection and Analysis

We interviewed participants using a semi-structured approach. The first part of the interview was not recorded and was intended to ensure advocates were comfortable talking to us. We explained the scope of the study, positioned tensions as inevitable and productive in RAI and in organizations more broadly, and addressed any confidentiality concerns. In addition to signing the IRB forms, many participants asked us to not disclose the name of their company, their formal role, or provide identifying details about the type of work they did. When participants were ready, we started the video recording and asked them to explain their work and how it related to RAI. Interview topics included examples of how they had advocated for RAI; examples of doing the right thing for RAI in spite of pressures; examples of push-back, if any, they had encountered; advice for others in the form of Dos and Don'ts of RAI advocacy; and ways they managed energy – both finding support when drained and staying energized.

Most interviews were typically conducted by two researchers, one serving as main interviewer and the other as note-taker and back-up. Interviews were scheduled for one hour, and they lasted 54 minutes on average. The longest interview was 69 minutes and the shortest, 29 minutes. In total, we collected about 20 hours of interview recordings, which were automatically transcribed by the qualitative data analysis software Marvin.

We read and corrected the automated transcripts, then coded sections of interviews by adding notes and labels in Marvin. Given the semi-structured, conversational nature of our interviews, we used an approach that combined inductive coding from interviews and consulting existing literature to develop an interpretive framework from the data (Tracy and Gist-Mackey 2024).

We did two rounds of coding. The first round involved open coding. The second author led this round. Open coding involved creating a list of advocates' individual actions, such as "sharing examples" and "using numbers," and then connecting different actions into codes to capture strategies such as "raising awareness" and "clearly communicating." The initial outcome was a list of 21 advocacy strategies, based on 184 interview excerpts where participants discussed examples of RAI advocacy. The first author then worked with the second author to complete open coding, leading to the final set of 14 strategies. (We merged some strategies with others due to strong ties - e.g., "find the lowest common denominator" was merged with "tie RAI work to business goals and product needs.")

The second round involved axial coding to code the interviews for organizational challenges. The first author led this round, working with second and third authors. We used interview data and existing literature to identify five broad challenges discussed by RAI advocates related to promoting RAI within their organizations. We then created second-level codes that linked strategies to challenges (Tracy 2018; Scott and Medaugh 2017). We observed that challenges provide an organizing framework for advocacy strategies, as participants adapted their strategy depending on the challenge they faced and the goal they aimed to achieve. In addition to challenges and advocacy strategies, our findings highlight the mental and emotional toll of RAI work, which is a separate theme.

4 Findings

We organize our findings around challenges advocates shared in interviews. These challenges are also supported by existing research literature, and they align with two major tensions: getting buy-in for RAI (the *why* of RAI) and facilitating RAI work (the *how* of RAI). For each challenge, we describe the common communicative strategies advocates referenced in addressing it -— how they offer new meanings in response to the underlying tensions. An advocacy strategy (e.g., "leverage existing beliefs and values") can be implemented via different actions (e.g., "position RAI as something people already do," "appeal to safety concerns," and "align work with other teams"). In this paper, we focus on 1-2 actions per strategy and feature only the most salient quotes. We end with findings about the emotional aspects and toll of RAI advocacy work as described by advocates during interviews.

4.1 Tension: Why? Getting Buy-In for RAI

Before any action toward RAI can take place, advocates have to get buy-in from peers and leaders. Participants mentioned two challenges for getting buy-in: lack of awareness about RAI and the inability to see how RAI is relevant to a specific case.

Challenge 1: Lack of RAI Awareness

Participants explained that stakeholders were at times simply unaware of RAI, a foundational challenge for advocacy. When addressing the lack of awareness, participants talked about using strategies such as *speaking up at the appropriate time, using examples as fear appeals* (Tannenbaum et al. 2015), and *cultivating audiences and adapting to them*.

Speaking up when peers and managers were discussing a topic with RAI implications was one of the most straightforward strategies for raising awareness. P18 explained that doing RAI advocacy in the real context of their job "*looks like raising my hand in the meeting and speaking to [RAI].*" Several pointed to meetings, readouts, or workshops as key occasions for advocacy, both because of the opportunity to get engagement on RAI from a wider audience and to “get everyone on the same page” (P02).

A related strategy to raise awareness was to ground RAI concerns in real world and at times personal *examples*. Some used hypothetical examples to draw attention to what they think are “*the riskiest scenarios*” (P03). Others drew from experiences or examples that were both familiar within their organizations and portrayed in media to create fear appeals (Tannenbaum et al. 2015):

> “Being able to talk about those [cases] and present them in a way that, hey, you know, we know this happened. It was embarrassing. This is how we took care of it. We want to walk you through it so you don’t end up in the same kind of situation.” (P12)

Pointing to existing examples or hypotheticals was common, and some advocates also drew examples from their own personal experiences, identities, or values. One advocate (P17) described how growing up in East Africa allowed them to recognize and address cultural gaps in their work that others may overlook. They found this particularly useful in their fairness-related work and starting a team focused on globally inclusive data practices:

> “There’s nothing like a good anecdote or story, something personal that really makes your message land. And if it comes from a really authentic place, like this is my lived experience, it does make an impact at times. So I have in some cases used that to my advantage.” (P17)

Another noted that their involvement in their Latina community outside of work energized their advocacy efforts in their organization, enabling them to:

> “[bring] something up that’s going to meet the needs of my community. And it’s going to meet the needs, hopefully, or how people think about considerations that will impact other communities like mine, who are underrepresented in these meetings or are often the last ones to be able to access some innovation.” (P18)

P11 described imbuing personal values and anecdotes as showing vulnerability, and that they cared about the issues beyond the boundaries of work. However, drawing from personal identities and anecdotes also opened advocates up to microaggressions and the weaponization of their identities against their goals. For example, P17 described a presentation with a company executive on an RAI goal that they advocated for early in their career. Despite working with their management line to refine the substantive and business-oriented arguments of their presentation, they found that the executive interpreted their presentation as claiming that the AI technology was racist, refuting the premise of the work entirely:

> “I was really approaching it from [the view that] our customers care about [RAI] and so we should care about this. What are we going to do about it? But he kept bringing it back to race and racism.” (P17)

Advocates also emphasized that *cultivating audiences and adapting to them* to raise awareness was important for pushing RAI initiatives forward. Common strategies for audience building were to escalate issues through chains of management to get successive buy-in at each level. Additionally, advocates framed RAI within messaging that already appealed to audiences, especially executives. P16 explained:

> “For example, let’s say we have a leader who is super motivated by customer feedback, right? Then what are the ways that we understand who is usually providing them with those insights, and then figure out the right ways to start having customer insight sessions that are more focused on RAI topics and then have that person who is typically the input there present that out with the right lens.” (P16)

P15 used the crawl-walk-run management framework as a rhetorical strategy to “*present the size and nature of the [RAI] gap in a way that executives are used to.*”

> ”[Showing a] crawl-walk-run table, and having 50 rows and all of them are red, and they’re all crawl and being like, ‘this is where we are now.’ And, I can explain to you why we’re at crawl in all of these. And then they’re like, ‘what? Okay, get them to walk.’ And I’ll say, okay, we can do that but in this timeline if we have this resourcing. And here’s exactly how I would do it with that many people.” (P15)

Challenge 2: RAI’s Relevance in Specific Cases

Beyond raising awareness, advocates described challenges in showing how RAI was relevant in specific situations. In many instances, others were aware of general RAI principles or examples but could not see how their existing knowledge about RAI was relevant in a specific setting. In these cases, advocates used strategies aimed at showing relevance and making abstract RAI issues concrete. The specific strategies were *showing the business value of RAI, leveraging existing policies and organizational practices,* and *leveraging existing beliefs and values related to software development.*

Advocates would illustrate *the business value* of RAI by showing how RAI goals and business goals are compatible or at least not in tension. Several participants talked about discussing RAI in the context of a specific product and showing how RAI contributes to business value and a better product overall. One way of advocating for RAI in products is through the perspective of user trust – that safe and responsible products would be appreciated by users and become a distinguishing feature in market competition:

> “[...] the point I’ve been trying to make for ages, is that actually, our users really care about the thought and effort and intention that’s being put into products. If I’m a user [...] of [company name] products, that is something that I would expect of the company building all of this technology. And so I’ve always seen that as, well, while people are saying that RAI is not always a business priority, people [do] care. That is a differentiator in the business anyway.” (P14)

In organizations with internal RAI policies, advocates would also *lean heavily on existing policies or practices* to make the case that RAI was relevant in a specific business context. Participants said that policies "*reduce the burden in terms of RAI teams like us going in and having to advocate for the work that we're doing*" (P05). Policies also provided a standard that advocates could point to: "*[RAI review committee] brings this new level of, like, okay, you've got to meet this minimum bar, or you cannot ship, right?*" (P16).

However, advocates did not have to always use policies as enforcement mechanisms, and many tended to avoid directly invoking policy when other strategies were on the table. For example, one adovcate used the metaphor of a hammer to describe the different ways of invoking policies:

> "Now, you as the RAI practitioner, you need to decide what you're going to do with that hammer [the policy]. Right? You can run around and then whack everything that's in your view with that hammer, you know, indiscriminately. Or you can go, oh, look, here's a nail. The first thing that I do with a nail usually is I don't whack it with full power. I actually hold it with my fingers, and I'm just going to tap it just a little bit to actually make sure that it sticks. And then hopefully, someone else is going to take it from there. I think it's all about what you do with the power that you have been given and the power, in this case, is policy and corporate strategy." (P11)

Even in organizations with low RAI maturity (Vorvoreanu et al. 2023; Heger et al. 2025), advocates could still *leverage existing values and beliefs*, such as the importance of privacy and security. For example, P08 highlighted specific product decisions with RAI implications to teams:

> "For a product decision, you decided not to run [AI] on confidential emails. Did you know that [that's] actually an RAI mitigation? Because now that's protecting privacy. [...] I think I have to map what they currently do. A lot of them think it's an addition. I'm like, it's not just addition, it's honestly what you already have decided and mapping it to the mitigation of these harms." (P08)

Several advocates emphasized that their organizations were far behind in RAI when compared with the maturity of security or privacy. Therefore, drawing parallels with existing attitudes toward security or privacy was a common and effective strategy for advocates when interacting with product teams.

4.2 Tension: How? Facilitating RAI Work

Even when others were bought into RAI, advocates frequently encountered challenges that hindered the progress of RAI work. The most common facilitation challenges that advocates discussed were overcoming adversarial perceptions of RAI and RAI advocates, uncertainty about how or why RAI concerns arise in specific contexts, and, at times, uncertainty about how to address RAI risks in practice.

Challenge 3: Overcoming Adversarial Perceptions of RAI and Advocates

At times, facilitating RAI work meant that advocates stepped into workstreams with existing timelines, scope, or goals made without considering RAI. By incorporating RAI objectives, advocates sometimes encountered resistance due to concerns that RAI objectives would compromise other aspects of the project. In situations where RAI was perceived as adversarial to the goals of a project, advocates focused on *fostering a partnership-oriented tone*, *highlighting goal alignment*, *strategically not using RAI language to make RAI happen*, or substituting *systems thinking* to walk through RAI concerns.

For example, one advocate described their early *tone-setting* with a product team that had never worked with their organization's RAI process as communicating:

> "I'm not here to block you. I'm really here to help you ship a more responsible product and that RAI is also going to be a distinguisher for you and the product that you ship." (P05)

The broader perception that RAI advocates were blocking or constraining the development of products was a challenge that advocates navigated with varying strategies. Some, like P07, sought to overcome the challenge by doubling down on how *RAI goals aligned with team goals*:

> "One quote I heard is 'if we do that, the RAI people will not be happy.' I thought, interesting, like, how can we become allies rather than trying to hide from one another. [...] In the end, it's building more sustainable relationships, more trustful relationships with customers." (P07)

Others emphasized *strategically not using RAI language to make RAI happen*:

> "[I refrain from] using the word RAI constantly with our product [team]. We have our principles. We're talking about good UX. We're framing it from the user perspective, and I think that helps." (P13)

Similarly, there were times when they would focus on "*systems thinking*," walking through risks and their potential implications with their audience without referencing RAI:

> "If we're sensing in the moment that the terms that represent things have already created a perception in a group of people that is not going to be in service of having them situated and attentive to the conversation that we want to run, push all of it aside then, and instead just speak to, 'Okay, so let's play this scenario out in a year's time, and let's imagine these three things happening. How could we mitigate the risk of that happening?' And then we're kind of getting people directly situated into thinking about systems, rather than labeling it as 'we're now going to do a RAI workshop' or 'we're now going to harms model these three scenarios.' That has been a tactic that I'm very conscious of and one that I've engaged in if I felt that actually the politics of the terms being used is taking up more airtime and is triggering

people to such an extent that they are more passionate about campaigning against doing it, than just getting on with the task at hand." (P09)

Challenge 4: Identifying RAI Concerns in Specific Contexts

Both advocates and stakeholders faced challenges related to identifying whether a particular product or feature had RAI concerns, and if so, deciding how to address them. To identify whether RAI concerns arose in products, advocates drew on strategies that *sought collaborative input from a range of expertise*, and focused on *communicating RAI information and concerns in concrete, focused formats* that matched the context of the products or teams.

Collaborative input was key to identifying RAI concerns in specific contexts. Advocates frequently described meeting or speaking with multiple stakeholders to collectively understand the implications of a product. P18 for instance, described meeting with team members to clarify and understand specific aspects of a product in order to build a clear picture of its RAI implications. Others, like P13, a UX researcher, described how they would make meetings about documenting information about an AI system's harms "*a collaborative activity, and then constantly repeating the sense-making that we came to together*."

Conveying RAI information in concrete, focused formats for others was important for fostering collaborative input. Advocates emphasized the importance of contextually relevant, tangible documents that could guide collective collaboration and decision-making. P14 emphasized the benefits of contextual relevance for teams, explaining:

> "When I point them towards a general resource, sometimes it can make my job a little bit harder, because then they come back to me with a lot of questions. So it sometimes makes more sense for me to spend an afternoon writing up a page that walks teams through this rather than trying to point them towards these company-wide resources." (P14)

Contextually relevant, tangible resources like short slide decks also enabled teams to raise concrete concerns through their management chains. P17 explained:

> "[It's one thing to say] as an organization, we're really failing in this area. And you can continue calling that out in the right meetings and in front of the right people. But if the right people don't have something to hold on to for them to advocate on your behalf, whether it's a one pager, a deck, a recording, something – they need a tangible thing that articulates the issue and what you're proposing, it's hard for them to carry your message forward." (P17)

Metrics also aided in describing RAI risks in concise formats and allowed teams to "*frame [RAI] goals in a more positive language*." P15 further explained:

> "We aim to reduce X amount and look at this long-term trend that we've achieved. Instead of just like, look at this. Imagine how bad it would have been if it happened." (P15)

However, advocates also raised concerns about a proliferation of metrics in the RAI space, pointing out that a focus on metrics could depersonalize harms or inadequately justify the sociotechnical problems in RAI. As P20 noted:

> "When I could write up research proposals, especially for research on inclusion and diversity or sustainability [for RAI], my boss would want to see an actual metric of, like, how much carbon are you removing through this research. Which is an impossible ask." (P20)

Challenge 5: Uncertainty About How to Address RAI Concerns in Products

Advocacy often occurred in situations where there was uncertainty about how to address RAI concerns. In these cases, advocates embraced pragmatism, meaning that they *focused on incremental wins and compromises that advanced some RAI goals*. They also described *relying on their own networks* to discuss and develop advocacy strategies and attempt to embed advocacy more broadly.

One way to reduce uncertainty about how to address RAI concerns was by *focusing on incremental wins and conveying pragmatism*. To convey pragmatism, some discussed problem-based framing of issues, focusing on shared goals or beliefs, and "the next best thing" (P11) in terms of RAI mitigations.

Advocates also explained that RAI was difficult to do alone. *Personal networks* were important for addressing RAI concerns in uncertain situations because they could draw from a larger knowledge base for help. Moreover, larger networks allowed RAI advocates to collectively make sense of new AI technologies. P10, for instance, explained how a community of practice within their company around RAI was crucial for understanding implications of generative AI:

> "We've built a bit of a community around RAI, and that was intentional as well. So when all the generative AI, like when the generative AI feature started to ship and we all had to start engaging in these processes together, it really worked to say, hey, we're all figuring this out together. [...] And then as we're figuring it out together, then you're getting that sharing of learnings and, and then that's when you get kind of the spreading out of the advocacy." (P10)

Several participants emphasized using networks to spread advocacy and as structures to embed advocacy in organizations:

> "It's impossible to do that if you don't have some kind of collective action, even if that is like, in an informal sense, [...] but you need to have, like, networks of people who are talking about it and strategizing about it at an enterprise-wide level." (P20)

Others saw networks as helping secure resources:

> "How do we pull in the right people who are already passionate about this work, have already started momentum on this work and gather those people and laser-focus our efforts in a way that could actually make impact, like bring in budget so we can actually do stuff or build stuff?" (P17)

Creating networks ultimately served the purpose of fostering collaboration and innovation on RAI mitigations, enabling individuals to leverage each other's strengths and expertise, as P02 points out:

> "This is something that people can come together over that maybe they wouldn't have had a reason to come together over before. Like, these are folks who never would have spoken to one another. And maybe in a one-on-one situation or a one-on-one scenario, they may benefit from connecting on a topic like RAI." (P02)

4.3 The Emotional Toll of RAI Advocacy Work

In addition to challenges and communicative strategies to address them, another finding pertains to the significant emotional and psychological tolls of RAI advocacy. P15 summarized several common themes related to the taxing nature of RAI advocacy:

> "[...] there's that tax on people. So there's the tax of trying to be proactive and doing this additional work, on a kind of volunteer basis, or putting together a plan. All that pre-work is taxing. The advocacy itself is taxing. Making resource requests and getting shut down over and over again is taxing. It hurts your self-esteem." (P15)

Advocacy work was frequently taken on voluntarily in addition to formal job duties. Advocates recognized that RAI work was needed or would not be done unless they stepped in, taking on the additional workload and thereby adding to the emotionally and psychologically draining nature of RAI work. Those who did RAI-related work as part of their formal role still faced frustration from under-resourcing or being turned down. Furthermore, a common theme was that advocates were drawn to RAI as a personal passion, amplifying the tolls of rejections and setbacks. Several described how RAI setbacks "*felt personal*" or "*hurts your self-esteem*," as P15 explained.

Many also noted that their organizations made ostensive commitments to RAI without providing additional support, which further contributed to the taxing feeling of RAI advocacy. P12, for example, described their frustration with hollow messaging from the leaders in their company:

> "I think the one thing that could have gone better for us is leadership support. And of course, our leaders are going to say, 'yes, yes, we love RAI.' But it just didn't have validation from an OKR [objectives and key results] standpoint, or people weren't really being recognized for it." (P12)

Similarly, P17 noted that advocacy work felt overwhelming, "*especially when you recognize sometimes that the company could put more dollars towards this [...but] just choose not to. I feel like that's when [...] you feel super deflated.*"

Several advocates remarked that the draining nature of RAI advocacy work led them to question whether they could sustain it long term. Some noted that they became more selective in how they advocated for RAI or whom they worked with on RAI advocacy. P04, for instance, reflected that they advocated for RAI before a range of people, from "allies to people that don't care, and more in-between" and "realized how draining it is to work with those that don't care [...and] since [their] time and capacity is limited anyway, [...they] mainly focused on those that care most of the time."

In summary, the emotional and psychological toll of RAI advocacy work stemmed not only from the inherent challenges of pushing for RAI practices but also from a lack of validation. The lack of formal recognition, insufficient resources, and perceived organizational indifference to meaningful RAI progress amplified feelings of frustration, deflation, and burnout. These challenges often forced advocates to weigh their passion for RAI against their personal well-being, leading some to recalibrate their efforts or reconsider their long-term commitment to advocacy.

5 Discussion

Our findings illuminate the invisible work that individuals within organizations invest to get RAI work done. Our interviews with RAI advocates at large technology companies revealed five primary challenges to RAI work, which are aligned with two core tensions: the *why* and the *how* of RAI work. The first two challenges, lack of awareness about RAI and unclear relevance of RAI, are related to addressing the *why* of RAI: Why does RAI matter? The remaining three challenges, perceptions of RAI advocates as adversaries, uncertainty about RAI concerns in specific products, and about how to address them, are related to the *how* of RAI: How do we do RAI?

RAI advocates addressed these challenges using sophisticated, adaptive communicative strategies that include appealing to each audience's interests, using relevant real-world examples, adapting the format and content of communication to the audience, and leveraging common business interests as well as policy, processes, and social networks. These everyday communicative strategies shape the ongoing organizing for RAI within the structural tensions in large technology companies. They demonstrate what "power-with" (Follett 1925; Berman and Van Buren 2015) can look like in organizations as individuals use communicative strategies to collaborate on solutions and foster collective ownership of RAI work. The RAI advocacy strategies we identified are themselves a practical contribution–a living artifact that others, especially advocates at different companies, may find helpful when navigating similar challenges.

RAI advocates invest significant cognitive and emotional load as they try to anticipate what would appeal to each of their audiences and adapt appeals accordingly while sometimes being perceived as adversaries. During interviews, participants' non-verbal communication signaled a mix of passion, exhaustion, frustration, and hope. Taking into consideration that none of them had formal training in negotiation, RAI advocacy takes tremendous effort, as it means stretching beyond their usual areas of expertise and extracts a significant psychological and emotional toll.

Overall, our findings show that RAI advocacy is here to stay but is unsustainable in its current form. The existence of formal RAI policies and processes does not make RAI advocacy obsolete. The success of RAI still largely depends on advocacy, even as organizations' RAI processes mature. To make RAI advocacy sustainable, companies must recognize this fact and reward and support RAI advocacy work. In this section, we expand on these two points and discuss their implications for RAI research and practice, in addition to outlining the limitations of our work.

5.1 The Value of RAI Advocacy

Although participants mentioned the existence of formal RAI policies, processes, and educational resources in their organizations, which are indicators of emerging and developing levels of RAI maturity (Vorvoreanu et al. 2023; Tabassi 2023; Heger et al. 2025), they also made clear that these are not enough. Until organizations reach high levels of maturity characterized by product team members knowing how to identify, measure, mitigate, and monitor RAI risks in various contexts, RAI advocacy work will still be needed. RAI policies can outline *what* principles and criteria AI systems must meet but cannot specify *how* to achieve these criteria in different scenarios. RAI advocates must connect the dots and help product teams see how RAI applies in specific contexts. Also, RAI work is not always straightforward. It often includes difficult trade-offs and bespoke solutions that require specialized expertise. RAI advocates facilitate RAI work by providing that expertise themselves or leveraging their networks. These are reasons to believe that RAI advocacy is not only here to stay but will remain a crucial component of organizational RAI efforts.

These findings on *how* and *why* RAI advocacy is crucial to the success of RAI help advance RAI scholarship on at least two fronts. The first front is research on RAI's sociotechnical and organizational challenges. Identifying and addressing such challenges are core concerns for RAI researchers (Whittlestone et al. 2019; Madaio et al. 2024, 2020; Ali et al. 2023; Passi and Barocas 2019; Rakova et al. 2021; Passi 2021). While research has highlighted a wide array of RAI challenges, efforts to address them remain limited in their effectiveness (Schiff et al. 2020; Holstein et al. 2019; Piorkowski et al. 2021; Nahar et al. 2022). This is, in part, because there is little research on how practitioners resolve such challenges on the ground. Our findings advance critical scholarship on the practical work of addressing RAI challenges by illuminating how RAI advocates *overcome* and, at times, *work around* such challenges using strategic communicative moves. In doing so, we also open new avenues of research into, for instance, what best practices for RAI advocacy look like in different organizational contexts and how communication of, for, and around RAI shapes RAI aspirations, concerns, and practice.

The second front for advancing RAI scholarship is research on RAI frameworks, guidelines, and tools. Over the years, researchers have proposed and developed several such artifacts (e.g., Madaio et al. 2020; Mitchell et al. 2019; Gebru et al. 2021; Aler Tubella et al. 2023; Boggust et al. 2023; Mccradden et al. 2023; Tang et al. 2024). At the same time, however, researchers continue to highlight the disconnect between the simplistic top-down approaches embedded in such artifacts and the messy realities of RAI's bottom-up practice (Passi and Barocas 2019; Madaio et al. 2020, 2024; Ali et al. 2023; Holstein et al. 2019; Nahar et al. 2022; Deng et al. 2022, 2023; Heger et al. 2022). Our research bridges the gap between theoretical research and practical implementation of RAI by highlighting the necessity of RAI advocacy for the uptake and success of RAI artifacts. We also hope this inspires further research into, for instance, a deeper exploration of on-the-ground dynamics between RAI artifacts and practitioners, as well as systematizing and scaling the communicative work of RAI advocacy.

5.2 Making RAI Advocacy Sustainable

RAI advocacy makes RAI possible, yet remains unacknowledged, unrewarded, and unsupported. In addition to complementing recent scholarship on how specific types of RAI work remain invisible in organizations (e.g., Deng et al. 2022; Passi and Barocas 2019; Widder et al. 2023), our findings highlight that RAI advocacy work is exhausting. Although critical to the success of organizational RAI efforts, RAI advocacy work remains unsustainable due to its high cognitive, psychological, and emotional toll. RAI advocacy work is complex. Our findings show that RAI advocacy requires, among other things, sophisticated communication and negotiation skills—skills that (R)AI practitioners often lack training in and researchers frequently overlook when examining on-the-ground RAI challenges. RAI's incentivization and resourcing challenges only make matters worse (Rakova et al. 2021; Heger, Passi, and Vorvoreanu 2022; Madaio et al. 2020). This results in situations where the success of organizational RAI efforts often hinges on the voluntary work of a handful of passionate individuals who relentlessly advocate for RAI with little to no help, recognition, or reward.

To make RAI advocacy more sustainable, organizations must first strive to make RAI advocacy work visible. RAI advocacy needs to be formally recognized, its necessity acknowledged, its complexities appreciated, and its work prioritized and rewarded. Our findings suggest that a *lot* of RAI advocacy work happens in technology companies, though it is rarely someone's actual job. This must also change. RAI advocacy needs to be recognized as a formal component of people's work, supported by organizational culture, practices, and structure. For instance, organizations must recognize the limitations of fixed roles for RAI and promote more exhaustive RAI advocacy efforts, embracing a "power-with" (Follett 1925; Berman and Van Buren 2015) approach to foster collective ownership of RAI work. Organizations should consider investing in supporting RAI advocates. For example, the communicative strategies this work has identified could be shared so RAI advocates can learn from each other what works and what doesn't. Organizations could also help foster–as several participants mentioned–an RAI community where they can support each other professionally and emotionally.

Ultimately, organizations should work to make RAI advocacy unnecessary. This could be achieved by investing in increasing organizational RAI maturity (Vorvoreanu et al. 2023; Heger et al. 2025). Formal, at-scale RAI advocacy in the form of education, training, and employee motivation, as well as the availability of varied resources product teams can draw upon as they learn to practice RAI, are steps towards higher RAI maturity. However, these are unlikely to happen without commitment and investment from top leadership. As one participant put it, "*I don't think C-suite gives a fuck, like, at all*". Whether perception or reality, this, too, needs to change.

5.3 Limitations

While this research takes a first step and provides valuable insights into the role of RAI advocates in large technology companies, it is not without limitations.

First, the research relies on individual self-reported data from the interviewees, which may be subject to biases such as social desirability or recall bias. Participants' accounts of their experiences and strategies may not always accurately reflect their actual practices or the broader organizational context and, like most social interactions, might be viewed differently from others' perspectives. Moreover, while we observed that participants shared their experiences candidly, it is possible that there were topics and individuals they were not comfortable discussing with us that they left out.

Second, the rapidly evolving nature of AI and RAI means that the findings of this study may become outdated as new technologies, policies, and organizational practices emerge. Future research should continue exploring the dynamic landscape of RAI advocacy. At the time the data for this study was collected, the AI policy climate in the United States was conducive to RAI. RAI advocacy might be different in changing policy climates, especially in companies eager to align with the political currents du jour.

6 Conclusion

This research highlights the critical role of RAI advocates in large technology companies. Through an analysis of 22 in-depth interviews, we identified five challenges these advocates face and a sophisticated set of communicative strategies they use to address them.

While not generalizable across the tech sector, these findings underscore the demanding and often under-recognized work of RAI advocacy. By examining the communicative strategies used by RAI advocates, this research sheds light on the invisible work that enables and shapes RAI efforts in organizations. We offer insights into why RAI advocacy is here to stay and how organizations can better support RAI advocates. Ultimately, this research envisions a future where RAI is seamlessly integrated into everyday organizational processes, reducing the need for advocacy.

Further studies could investigate advocacy strategies in various organizational, cultural, and policy contexts, and look into their effectiveness. Future research could also benefit from borrowing from organizational communication the theoretical lens of communication as constitutive of organizations (CCO) and critically examine how work, organizations, policy, values, and more are enacted through mundane communicative acts. Advocacy and resistance in organizations can take many forms, and might not always be formal or even highly visible. As an Ethiopian proverb goes, "When the great lord passes, the wise peasant bows deeply and silently farts." The CCO perspective enables us to notice, understand, and appreciate the small acts of resistance, the daily tensions through which we continuously constitute our social realities.

Ethical Statement

As an internal research team at a large technology company, we recognized that the subject of this research itself poses ethical challenges, both due to the sensitive and personal nature of participants' experiences with RAI advocacy and the potential for participants' professional reputational harm. Beyond obtaining IRB approval for original human-subjects research and following informed consent processes for this paper, we made several considerations to affirm the dignity, privacy and confidentiality of participants. We collected a limited amount of participant demographic and biographic information, and only report aggregated counts. We intentionally sought perspectives from outside of our organization, as well as diversity in the identities, professional roles, and demographics of participants. We also offered participants an opportunity to attend a presentation of preliminary findings.

Acknowledgments

We wish to thank our research participant for sharing their RAI advocacy experiences and strategies with us. We also wish to thank the following individuals for help with this research: Shipi Dhanorkar, Sydney Graham, Amy K. Heger, Dennis Mumby, and Kathleen Walker. Last but not the least, we thank the anonymous reviewers for their feedback.

References

Aler Tubella, A.; Coelho Mollo, D.; Dahlgren Lindström, A.; Devinney, H.; Dignum, V.; Ericson, P.; Jonsson, A.; Kampik, T.; Lenaerts, T.; Mendez, J. A.; and Nieves, J. C. 2023. ACROCPoLis: A Descriptive Framework for Making Sense of Fairness. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '23, 1014–1025. New York, NY, USA: Association for Computing Machinery. ISBN 9798400701924.

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*, FAccT '23, 217–226. New York, NY, USA: Association for Computing Machinery. ISBN 9798400701924.

Ashcraft, K. L.; Kuhn, T. R.; and Cooren, F. 2009. 1 Constitutional Amendments: "Materializing" Organizational Communication. *Academy of Management Annals*, 3(1): 1–64.

Berman, S. L.; and Van Buren, H. J. 2015. Mary Parker Follett, managerial responsibility, and the future of capitalism. *Futures*, 68: 44–56.

Bisel, R. S. 2018. *Organizational moral learning: a communication approach*. New York, NY: Routledge, Thaylor & Francis Group. ISBN 978-1-138-11954-3 978-1-138-11956-7.

Bisel, R. S.; Kelley, K. M.; Ploeger, N. A.; and Messersmith, J. 2011. Workers' Moral Mum Effect: On Facework and Unethical Behavior in the Workplace. *Communication Studies*, 62(2): 153–170.

Boggust, A.; Suresh, H.; Strobelt, H.; Guttag, J.; and Satyanarayan, A. 2023. Saliency Cards: A Framework to Characterize and Compare Saliency Methods. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '23, 285–296. New York, NY, USA: Association for Computing Machinery. ISBN 9798400701924.

Browne, J.; Drage, E.; and McInerney, K. 2024. Tech workers' perspectives on ethical issues in AI development: Foregrounding feminist approaches. *Big Data & Society*, 11(1): 20539517231221780.

Brummans, B. H. J. M.; Cooren, F.; Dion, F.; and Vásquez, C. 2024. Chapter 2 Emerging Trends and Topics in Research on the Communicative Constitution of Organizations. In Miller, V. D.; and Poole, M. S., eds., *Organizational Communication Theory and Research*, 27–44. De Gruyter. ISBN 978-3-11-071850-8.

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.

Deng, W. H.; Yildirim, N.; Chang, M.; Eslami, M.; Holstein, K.; and Madaio, M. 2023. Investigating Practices and Opportunities for Cross-functional Collaboration around AI Fairness in Industry Practice. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '23, 705–716. New York, NY, USA: Association for Computing Machinery. ISBN 9798400701924.

Detert, J. R.; and Treviño, L. K. 2010. Speaking Up to Higher-Ups: How Supervisors and Skip-Level Leaders Influence Employee Voice. *Organization Science*, 21(1): 249–270.

Dutton, J. E.; Ashford, S. J.; O'Neill, R. M.; and Lawrence, K. A. 2001. MOVES THAT MATTER: ISSUE SELLING AND ORGANIZATIONAL CHANGE. *Academy of Management Journal*, 44(4): 716–736.

Follett, M. 1925. Power. Public lecture reprinted (1995). *Mary Parker Follett, prophet of management: A celebration of writings from the 1920s*, 97À119.

Garner, J. T. 2013. Dissenters, Managers, and Coworkers: The Process of Co-Constructing Organizational Dissent and Dissent Effectiveness. *Management Communication Quarterly*, 27(3): 373–395.

Garner, J. T.; and Garner, L. T. 2011. Volunteering an Opinion: Organizational Voice and Volunteer Retention in Nonprofit Organizations. *Nonprofit and Voluntary Sector Quarterly*, 40(5): 813–828.

Garner, J. T.; and Ragland, J. P. 2019. Tabling, discussing, and giving in: Dissent in workgroups. *Group Dynamics: Theory, Research, and Practice*, 23(1): 57–74.

Gebru, T.; Morgenstern, J.; Vecchione, B.; Vaughan, J. W.; Wallach, H.; III, H. D.; and Crawford, K. 2021. Datasheets for datasets. *Commun. ACM*, 64(12): 86–92.

Hatton, E. 2017. Mechanisms of invisibility: rethinking the concept of invisible work. *Work, Employment and Society*, 31(2): 336–351.

Heger, A.; Passi, S.; and Vorvoreanu, M. 2022. All the tools, none of the motivation: organizational culture and barriers to responsible AI work. *Cultures in AI/AI in Culture: a NeurIPS 2022 workshop.*

Heger, A. K.; Marquis, L. B.; Vorvoreanu, M.; Wallach, H.; and Wortman Vaughan, J. 2022. Understanding Machine Learning Practitioners' Data Documentation Perceptions, Needs, Challenges, and Desiderata. *Proc. ACM Hum.-Comput. Interact.*, 6(CSCW2).

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).

Hegstrom, T. G. 1990. Mimetic and dissent conditions in organizational rhetoric. *Journal of Applied Communication Research*, 18(2): 141–152.

Holstein, K.; Wortman Vaughan, J.; Daumé, H.; Dudik, M.; and Wallach, H. 2019. Improving Fairness in Machine Learning Systems: What Do Industry Practitioners Need? In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, 1–16. New York, NY, USA: Association for Computing Machinery. ISBN 9781450359702.

Jakesch, M.; Buçinca, Z.; Amershi, S.; and Olteanu, A. 2022. How Different Groups Prioritize Ethical Values for Responsible AI. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '22, 310–323. New York, NY, USA: Association for Computing Machinery. ISBN 9781450393522.

Jing, F. S.; Berger, S. E.; Becerra Sandoval, J. C.; Pepper, K.; Wheeler, A. M.; Mayoral, P. R.; Lokesh, D.; Feng, A.; Mijalkovic, M.; Bao, C.; Dholakia, S.; and Goyal, M. 2024. Designing for Agonism: 12 Workers' Perspectives on Contesting Technology Futures. *Proc. ACM Hum.-Comput. Interact.*, 8(CSCW1).

Kajzer Mitchell, I.; and Walinga, J. 2017. The creative imperative: The role of creativity, creative problem solving and insight as key drivers for sustainability. *Journal of Cleaner Production*, 140: 1872–1884.

Kassing, J. W. 2002. Speaking Up: Identifying Employees' Upward Dissent Strategies. *Management Communication Quarterly*, 16(2): 187–209.

Kundro, T. G.; Neely, B. H.; and Muir, C. P. 2024. Supervisor integrity empowers employees to advocate for diversity in problematic climates. *Journal of Applied Psychology*, 109(5): 668–686.

Leonardi, P. M. 2011. Innovation Blindness: Culture, Frames, and Cross-Boundary Problem Construction in the Development of New Technology Concepts. *Organization Science*, 22(2): 347–369.

Leonardi, P. M. 2015. Studying Work Practices in Organizations: Theoretical Considerations and Empirical Guidelines. *Annals of the International Communication Association*, 39(1): 235–273.

Lu, S.; Bartol, K. M.; Venkataramani, V.; Zheng, X.; and Liu, X. 2019. Pitching Novel Ideas to the Boss: The Interactive Effects of Employees' Idea Enactment and Influence Tactics on Creativity Assessment and Implementation. *Academy of Management Journal*, 62(2): 579–606.

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.

Mazmanian, M.; Cohn, M.; and Dourish, P. 2014. Dynamic Reconfiguration in Planetary Exploration: A Sociomaterial Ethnography. *MIS Quarterly*, 38(3): 831–848.

Mccradden, M.; Odusi, O.; Joshi, S.; Akrout, I.; Ndlovu, K.; Glocker, B.; Maicas, G.; Liu, X.; Mazwi, M.; Garnett, T.; Oakden-Rayner, L.; Alfred, M.; Sihlahla, I.; Shafei, O.; and Goldenberg, A. 2023. What's fair is… fair? Presenting JustEFAB, an ethical framework for operationalizing medical ethics and social justice in the integration of clinical machine learning: JustEFAB. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '23, 1505–1519. New York, NY, USA: Association for Computing Machinery. ISBN 9798400701924.

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*, FAT* '19, 220–229. New York, NY, USA: Association for Computing Machinery. ISBN 9781450361255.

Mittelstadt, B. 2019. Principles alone cannot guarantee ethical AI. *Nature machine intelligence*, 1(11): 501–507.

Mumby, D. K. 2005. Theorizing Resistance in Organization Studies: A Dialectical Approach. *Management Communication Quarterly*, 19(1): 19–44.

Mumby, D. K.; and Kuhn, T. R. 2024. MARY PARKER FOLLETT: BRIDGING THEORY AND PRACTICE. In *Organizational Communication: A Critical Introduction*. Los Angeles, Calif: SAGE Press, 3rd edition.

Mumby, D. K.; Thomas, R.; Martí, I.; and Seidl, D. 2017. Resistance Redux. *Organization Studies*, 38(9): 1157–1183.

Nahar, N.; Zhou, S.; Lewis, G.; and Kästner, C. 2022. Collaboration challenges in building ML-enabled systems: communication, documentation, engineering, and process. In *Proceedings of the 44th International Conference on Software Engineering*, ICSE '22, 413–425. New York, NY, USA: Association for Computing Machinery. ISBN 9781450392211.

Nedzhvetskaya, N.; and Tan, J. S. 2024. The Role of Workers in AI Ethics and Governance. In *The Oxford Handbook of AI Governance*. Oxford University Press. ISBN 9780197579329.

Orlikowski, W. J.; and Gash, D. C. 1994. Technological frames: making sense of information technology in organizations. *ACM Transactions on Information Systems*, 12(2): 174–207.

Orlikowski, W. J.; and Scott, S. V. 2015. Exploring Material-Discursive Practices. *Journal of Management Studies*, 52(5): 697–705.

Passi, S. 2021. *Making data work: the human and organizational lifeworlds of data science practices*. Ann Arbor, Michigan: ProQuest Information and Learning.

Passi, S.; and Barocas, S. 2019. Problem Formulation and Fairness. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, FAT* '19, 39–48. New York, NY, USA: Association for Computing Machinery. ISBN 9781450361255.

Piderit, S. K.; and Ashford, S. J. 2003. Breaking Silence: Tactical Choices Women Managers Make in Speaking Up About Gender-Equity Issues*. *Journal of Management Studies*, 40(6): 1477–1502.

Piorkowski, D.; Park, S.; Wang, A. Y.; Wang, D.; Muller, M.; and Portnoy, F. 2021. How AI Developers Overcome Communication Challenges in a Multidisciplinary Team: A Case Study. *Proc. ACM Hum.-Comput. Interact.*, 5(CSCW1).

Putnam, L. L.; Fairhurst, G. T.; and Banghart, S. 2016. Contradictions, Dialectics, and Paradoxes in Organizations: A Constitutive Approach. *Academy of Management Annals*, 10(1): 65–171.

Rakova, B.; Yang, J.; Cramer, H.; and Chowdhury, R. 2021. Where Responsible AI meets Reality: Practitioner Perspectives on Enablers for Shifting Organizational Practices. *Proc. ACM Hum.-Comput. Interact.*, 5(CSCW1).

Satterstrom, P.; Kerrissey, M.; and DiBenigno, J. 2021. The Voice Cultivation Process: How Team Members Can Help Upward Voice Live on to Implementation. *Administrative Science Quarterly*, 66(2): 380–425.

Schiff, D.; Rakova, B.; Ayesh, A.; Fanti, A.; and Lennon, M. 2020. Principles to Practices for Responsible AI: Closing the Gap. arXiv:2006.04707.

Schoeneborn, D.; Kuhn, T. R.; and Kärreman, D. 2019. The Communicative Constitution of Organization, Organizing, and Organizationality. *Organization Studies*, 40(4): 475–496.

Scott, C.; and Medaugh, M. 2017. Axial Coding. In Matthes, J.; Davis, C. S.; and Potter, R. F., eds., *The International Encyclopedia of Communication Research Methods*, 1–2. Wiley, 1 edition. ISBN 978-1-118-90176-2 978-1-118-90173-1.

Star, S. L.; and Strauss, A. 1999. Layers of Silence, Arenas of Voice: The Ecology of Visible and Invisible Work. *Computer Supported Cooperative Work (CSCW)*, 8(1): 9–30.

Tabassi, E. 2023. Artificial Intelligence Risk Management Framework (AI RMF 1.0).

Tang, N.; Zhi, J.; Kuo, T.-S.; Kainaroi, C.; Northup, J. J.; Holstein, K.; Zhu, H.; Heidari, H.; and Shen, H. 2024. AI Failure Cards: Understanding and Supporting Grassroots Efforts to Mitigate AI Failures in Homeless Services. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '24, 713–732. New York, NY, USA: Association for Computing Machinery. ISBN 9798400704505.

Tannenbaum, M. B.; Hepler, J.; Zimmerman, R. S.; Saul, L.; Jacobs, S.; Wilson, K.; and Albarracín, D. 2015. Appealing to fear: A meta-analysis of fear appeal effectiveness and theories. *Psychological bulletin*, 141(6): 1178.

Tracy, S. J. 2018. A Phronetic Iterative Approach to Data Analysis in Qualitative Research. *Journal of Qualitative Research*, 19(2): 61–76.

Tracy, S. J.; and Gist-Mackey, A. N. 2024. Chapter 5 Qualitative Methods in Organizational Communication. In Miller, V. D.; and Poole, M. S., eds., *Organizational Communication Theory and Research*, 85–104. Berlin, Boston: De Gruyter Mouton. ISBN 978-3-11-071850-8.

Vorvoreanu, M.; Heger, A.; Passi, S.; Dhanorkar, S.; Kahn, Z.; and Wang, R. 2023. Responsible AI Maturity Model. Technical Report MSR-TR-2023-26, Microsoft.

Weick, K. E. 1979. *The social psychology of organizing*. Topics in social psychology. New York: Random House, 2. ed edition. ISBN 978-0-394-34827-8.

Weick, K. E.; Sutcliffe, K. M.; and Obstfeld, D. 2005. Organizing and the Process of Sensemaking. *Organization Science*, 16(4): 409–421.

Whittlestone, J.; Nyrup, R.; Alexandrova, A.; and Cave, S. 2019. The Role and Limits of Principles in AI Ethics: Towards a Focus on Tensions. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '19, 195–200. New York, NY, USA: Association for Computing Machinery. ISBN 9781450363242.

Widder, D. G. 2024. Epistemic Power in AI Ethics Labor: Legitimizing Located Complaints. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*, 1295–1304. Rio de Janeiro Brazil: ACM. ISBN 9798400704505.

Widder, D. G.; Zhen, D.; Dabbish, L.; and Herbsleb, J. 2023. It's about power: What ethical concerns do software engineers have, and what do they (feel they can) do about them? In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '23, 467–479. New York, NY, USA: Association for Computing Machinery. ISBN 9798400701924.

Zanin, A. C.; Bisel, R. S.; and Adame, E. A. 2016. Supervisor Moral Talk Contagion and Trust-in-Supervisor: Mitigating the Workplace Moral Mum Effect. *Management Communication Quarterly*, 30(2): 147–163.