Who Should Act? Distancing and Vulnerability in Technology Practitioners' Accounts of Ethical Responsibility

KRISTINA POPOVA, Media Technology and Interaction Design, KTH Royal Institute of Technology, Sweden

CLÀUDIA FIGUERAS, Department of Computer and Systems Sciences, Stockholm University, Sweden

KRISTINA HÖÖK, Media Technology and Interaction Design, KTH Royal Institute of Technology, Sweden

AIRI LAMPINEN, Department of Computer and Systems Sciences, Stockholm University, Sweden

Attending to *emotion* can shed light on why recognizing an ethical issue and taking responsibility for it can be so demanding. To examine emotions related to taking or not taking responsibility for ethical action, we conducted a semi-structured interview study with 23 individuals working in interaction design and developing AI systems in Scandinavian countries. Through a thematic analysis of how participants attribute ethical responsibility, we identify three *ethical stances*, that is, discursive approaches to answering the question 'who should act': an individualised I-stance ("the responsibility is mine"), a collective we-stance ("the responsibility is ours"), and a distanced they-stance ("the responsibility is someone else's"). Further, we introduce the concepts of *distancing* and *vulnerability* to analyze the emotion work that these three ethical stances place on technology practitioners in situations of low- and high-scale technology development, where they have more or less control over the outcomes of their work. We show how the we- and they-stances let technology practitioners *distance* themselves from the results of their activity, while the I-stance makes them more *vulnerable* to emotional and material risks. By illustrating the emotional dimensions involved in recognizing ethical issues and embracing responsibility, our study contributes to the field of Ethics in Practice. We argue that emotions play a pivotal role in technology practitioners' decision-making process, influencing their choices to either take action or refrain from doing so.

CCS Concepts: • **Human-centered computing** → **Empirical studies in interaction design**.

Additional Key Words and Phrases: ethics, emotion, ethical stance, vulnerability, distancing, responsibility, ethics in practice

1 INTRODUCTION

Scholarly conversations about ethics have traditionally excluded emotion. Instead, moral philosophers have worked on defining the rational ground for ethical behaviour. This approach has been challenged by *ethics of care*, where theorists have demonstrated the situated nature of ethics and the role that feelings play in performing ethical action [72]. In the *phenomenology of ethics*, Varela

Authors' addresses: Kristina Popova, kpopova@kth.se, Media Technology and Interaction Design, KTH Royal Institute of Technology, Stockholm, Sweden, ; Clàudia Figueras, claudia@dsv.su.se, Department of Computer and Systems Sciences, Stockholm University, Stockholm, Sweden; Kristina Höök, khook@kth.se, Media Technology and Interaction Design, KTH Royal Institute of Technology, Stockholm, Sweden; Airi Lampinen, airi@dsv.su.se, Department of Computer and Systems Sciences, Stockholm University, Stockholm, Sweden.

challenges the understanding of ethics as a matter of rational judgment, demonstrating the role of embodied and, therefore, *felt* ethical reasoning [103]. Similarly, in the field of technology production, ethics has long been approached through rules and guidelines as attempts to avoid the biases of individual judgements that draw upon emotions and gut feelings. Recently, though, the role of emotion has been re-centered in approaches such as experience-centered design [4, 39]. Publications that examine the ethical concerns of technology practitioners empirically have also demonstrated the importance of emotions in both recognizing an ethical issue and acting upon it [99, 107]. We continue this exploration by looking at the connection between discourses regarding ethical responsibility and the feelings they generate among technology practitioners. We argue that attending to emotion can shed light on why taking ethical responsibility can be so demanding for technology practitioners. This is why feelings need to be considered in discussions of ethics, alongside principles like fairness and transparency.

Building on previous work that points at the importance of emotion in recognizing ethical concerns [107], we ask: What emotion work [4] does ethical responsibility place on technology practitioners? By 'emotion work,' we understand the management and navigation of emotional processes in work settings [49, 51]. Emotion work needs to be done when we, for example, handle emotions that conflict with the *feeling rules* of the workplace [49], that is, the established ways of demonstrating emotions. Such work is common when it comes to advancing ethical designs [4, 50, 95] or dealing with ethical contention in situations of value conflict [112].

We follow the tradition of studies of *Ethics in Practice* [13, 19, 42, 64, 65, 79, 91] by approaching the attribution of ethical responsibility from technology practitioners' perspective. We present an empirical study based on two independently collected sets of semi-structured interviews with altogether 23 practitioners in various sectors, from interaction designers working in academia to AI practitioners working for government agencies. Here, we focus on the affective dimension, that is, the emotions and feelings produced by different stances. We relate our exploration to studies of ethics in action in design and HCI literature [33, 48, 95] together with the studies focused on the emotion work of design researchers [4, 25, 50, 55, 82].

Our analysis is structured into three parts: First, through thematic analysis of the interview data, we identify three ethical stances – I, We, They – which indicate different approaches to attributing ethical responsibility. The I-stance ascribes responsibility directly to an individual, granting them a possibility to own the outcomes of an action but also putting them into a vulnerable position. The we-stance ascribes responsibility to a collective without specifying the role of an individual's action. The they-stance externalizes ethical responsibility, attributing it to a group or entity that the technology practitioner is not a part of. The stances should be understood as discourses: they are systematic ways of creating an ethical subject rather than individual psychological dispositions. Second, we analyze the differences between the three stances through the paired concepts of *distancing* and *vulnerability*. Each stance presumes a different degree of emotional engagement and personal involvement in the outcomes of the practitioner's work – engagement that makes a practitioner vulnerable to the risks related to both action and inaction. This can include emotional and material consequences. Third, we analyze the work that the three stances do in different types of environments and what consequences they may have for practitioners and the development of technology. We distinguish between *low-scale* and *high-scale* settings, that is, small vertically integrated design projects and technology development processes that rely on complex interdependencies of services and often go across organizations.

We make an empirical contribution by articulating three ethical stances, generated from our interviews with a diverse set of technology practitioners. Further, we introduce the concepts of *distancing* and *vulnerability* to analyze the emotion work that the ethical stances place on technology practitioners in situations of low- and high-scale technology development. Noting that

none of the stances is unproblematic, we articulate their emotional consequences: We argue that both over-individualizing ethical responsibility for the outcomes of complex coordinated activities and distancing oneself from acting ethically can become problematic in terms of the emotional distress and moral burden they place on technology practitioners. Here, we connect with scholars of *Ethics in Practice* [83, 92, 110] who argue for a more systemic approach to ethics that avoids placing too much pressure on individual technology practitioners. We conclude by discussing how we might draw upon resources from first-person design methods [52–54, 96] and research strategies from Science and Technology Studies [23, 28, 41, 61, 101] to shape conversations on ethics in a way that avoids over-individualizing ethical responsibility and alienating practitioners from the outcomes of their work.

2 BACKGROUND

Attributing responsibility for ethical action is not straightforward, especially not when it comes to the development of complex technical systems with multiple organizations and collaborators involved [19, 42, 44, 65, 91]. Even in situations where there is consensus on what it means to act ethically, the question of *who should act* often remains unresolved: an individual reflective designer, an organization following legislative and ethical guidelines, trained ethicists, social movements, or citizens? We draw inspiration from *Social Movements Studies* that analyze conditions for successful collective action for social change. According to the frame theory of social movements [9], providing a coherent narrative explaining what is wrong, what should be done to make things better, and who should act to implement the necessary changes is a key success factor. If social initiatives fail to articulate a motivating and comprehensive answer to the question of who should act, reasons for participation become unclear [9]. In order to demonstrate the parallels between the work that social movements do to motivate their supporters and the work done to promote ethical development of technology, we attend to three dimensions of scholarly debates regarding ethics in technology design and development: (1) introducing new values; (2) translating new values into practice, and; (3) attributing responsibility for taking ethical action. We demonstrate that the studies of ethics in design and AI do similar work by formulating problems, outlining solutions, and attributing responsibility. We centre our attention on the last dimension, attributing responsibility, by studying answers to the question *who should act*. Finally, we provide a review of prior empirical studies of ethics in practice and emotion work in relation to ethical work practice – the line of literature that most closely resonates with our research.

2.1 Which values? Introducing new ethical frameworks

The first dimension to consider is which values are promoted — or even seen as relevant enough to take into account. A key line of work to consider is *Design Ethics*, by which we mean efforts to make ethics tangible and applicable to the process of technology development (with approaches including but not limited to Value-Sensitive Design [34], Values at Play [32], Reflective Design [90], and Critical Design [6]); guiding for design [22, 63]; and visible in the details of the design process [33, 43, 91, 95]. Within Design Ethics, the work of introducing new values is most noticeable in the articulation of new agendas, such as ethics of care [24], feminist Human-Computer Interaction (HCI) [7], autonomous design [29], or design justice [22]. For example, in their article *Designing for existential crisis*, Light and Shklovski [63] suggest a new set of values and directions: designing for dignity instead of fairness, compassion instead of formalized rules, and re-imagining responsibility rather than empowering. They also urge readers to abandon other values and ways of thinking typical for HCI, such as separating between a designer and a user, designing for ease, and excluding non-human actors. This is an attempt to set up a new ethical agenda. Once such an agenda has been accepted or it has gained sufficient support so that it can be advanced, the emphasis can shift

to *how* values can be translated into practice. This entails a practice-oriented approach to how values come to matter within design processes and technology deployment.

The question of which values are given priority is also central in the AI Ethics literature[1]. We understand *AI Ethics* as a subfield of digital ethics that addresses political, social, and psychological implications of AI, including machine learning, big data analytics, and blockchain technologies [59], and seeks to mitigate or pre-empt harms that AI systems can cause [2, 16, 74]. *AI Ethics* advocates for developing AI that is trustworthy [1], ethical [75], or transparent, explainable, and accountable [106]. There is a lot of work within AI Ethics that sets out high-level ethical principles, such as privacy, fairness, and transparency [56] or responsibility and trustworthiness [1, 27, 75], to create guidelines for AI development teams.

2.2 How to implement new agendas? Translating values into practice

There have been numerous attempts at making values impactful and actionable for designers through creating toolkits and methodologies for ethics in design. For example, Value-Sensitive Design [34, 35] was developed to make the values of both designers and stakeholders part of the design process. A lot of attention has been paid to describing *how* to design ethically: how to visualise values, reconcile conflicting values, or identify stakeholders. After the introduction of the initial framework, researchers have applied Value-Sensitive Design to different areas of practice [5, 36, 38], evaluated the outcomes [37], and developed scalable toolkits [110] intended for different technological domains, including AI [111].

Within *AI Ethics*, the principle-based approach has been criticised for claiming universal applicability [45] and for not providing mechanisms to enact the principles [67]. The critique has pointed to the lack of guidance on navigating value tensions in practice [31, 45, 65] and framing ethics as a matter of expert oversight, where ethics is approached solely as a matter of technical or legal expertise rather than political discussion [45].

Answering the call for more actionable mechanisms for AI ethics (transitioning to an ethical *how* instead of the *what* [70]), technical teams have built AI ethics toolkits for implementing values in AI systems [60]. Most frequently, such toolkits concern fairness, bias mitigation, anti-discrimination, privacy, explainability, or accountability [46]. Many of the resulting toolkits are aimed at individual practitioners [110], which has led to criticism that the toolkits ignore the potential for collective action [110], as well as the role of organizational factors in ethical decision-making [84]. Another line of criticism relates to the claims that these tools are also limited in their attention to structural inequality [10, 26, 68] and the need for greater participation of marginalized, impacted communities, and the Global South, at large [26, 58, 81, 86].

2.3 Who should act? Attributing responsibility for ethical action

For us, the central question for ethics in technology design and development is *who should act*? Calls for designers to shoulder more moral responsibility are common both in scholarly discussion [57, 64, 69, 71] and popular design literature [69, 76]. This has led to efforts to make ethics part of design education, training reflexivity, and bringing new ethical perspectives into education [30]. For example, Sabie and Parikh [85] report on their experience of teaching students to engage with the communities they are designing for with care and engaging with values beyond productivity. In a recent systematic review of ethical studies in design, attributing responsibility for ethical action to a designer is shown to be the most common approach in Design Ethics [73]. There is a similar tendency in *AI Ethics* to place the responsibility on individuals, which has taken the form

---

[1]Our attention to AI, here, relates to the fact that half of our interviewees work in the domain of AI development.

of providing toolkits for developers [65, 66, 83] or stating the importance of ethical education to engineers and computer science students [13, 30].

Attributing ethical responsibility for technology development is complicated by the fact that responsibility is often distributed. For instance, since AI systems involve many actors working on production, data collection, and data mining, to name just a few, the distribution of responsibility is complex and related to *the problem of many hands* [102]. Coeckelbergh [21] points out that collective responsibility may mean that either a collective agent, such as an organization, is held responsible [77, 84, 88]) or that responsibility is distributed over a group of individual agents, each one bearing responsibility independently. In any case, individual responsibility may be considered less relevant in the context of digital technologies [87] because of the complex relationships among diverse actors, spaces and organizations. This resonates well with empirical studies focused on ethical complexity in organizations that demonstrate how organizational dynamics can prevent the fulfilment of individual good intentions [19, 65, 83, 92, 109, 110]. Not all long-term consequences can be accessed by an individual designer or taken into account within a complex organizational context [42]. Because of this, many studies conclude with recommendations to develop organizational procedures for implementing ethical design rather than placing the responsibility with the individual [42, 85], thus shifting the focus from the individual to organizational practice.

2.4 Empirical studies of ethics in practice

Research on ethics in practice studies how ethics *is done* as an integral part of the design process [33, 48, 82, 95]. For example, Shilton conducted an ethnography to analyze how software developers balance functionality and privacy requirements in designing new technology [91]. Together with Shilton, Boyd developed the concept of *ethical sensitivity* to distinguish between three analytical steps of implementing ethics in practice: noticing an ethical problem, understanding the situation, and taking a decision [13]. Within Participatory Design, Spiel showed how everyday work with participants, in their case children with autism, is an ongoing, processual, ethical matter [94, 95].

A recent line of studies has focused on the emotion work of design researchers who work in close proximity with their research participants, drawing attention to the traditionally hidden processes of managing emotions in research work [4]. Subsequent studies have drawn explicit attention to the relation between emotion work and ethics in design research [39, 50, 82], breaking the artificial dichotomy between *felt* experiences and rational behaviour. Considering the technology industry, Su and colleagues [99] conducted a study of *critical affects* among workers. Using the concept of *emotion habitus*, they analyzed how technology practitioners react to public criticisms of their field, removing the usual separation between ethics and emotions. Similarly, Widder and colleagues [107] demonstrated empirically the high psychological cost of raising ethical concerns within companies that develop AI products. Their work points to the importance of focusing the discussion on the distribution of power to solve ethical issues rather than just noticing them.

Scaling up and out from the specifics of individual participants in research settings, there is a growing number of empirical studies examining how ethics is enacted both in academia [105] and industry [19, 42, 64, 79, 109, 110]. One of the critical conclusions from this line of research is that organizational practices matter. While hardly a surprise to CSCW scholars, it is an observation with the potential to shape how we approach ethical responsibility in the organizational and collaborative settings where technology development takes place. Even in situations where an individual designer adheres to high moral standards, it is not guaranteed that deceptive design will be avoided [18, 19, 43, 110]. Learning how to navigate ethical complexities in organizations, then, is as important as adhering to the right set of values [19]. This practice-based view of ethics in organizations provides footholds for approaching ethical reasoning as an empirical matter and a part

of work practice where individuals need to reconcile their own assessment over what constitutes ethically justified action with the rules and structures constraining them.

3 RESEARCH MATERIALS AND METHODOLOGICAL APPROACH

We present an empirical study of how technology practitioners approach the question of *who should act?* Before presenting our materials and methods, we reflect on our positionality and explain the theoretical assumptions about ethics that guided our research.

3.1 Positionality and conceptual approach to ethics

We are a team working in academic institutions in Northern Europe, with diverse positions (background details anonymised for review). Importantly for this paper, we all work in close proximity to Interaction Design and the AI community. Our approach to ethical responsibility is contextualised by our own experiences of working predominantly in small teams but in overlaps with industry practitioners. Our diverse disciplinary backgrounds and our commitment to interdisciplinary research projects, from the social sciences to AI/ML, have made us more accepting of a diversity of approaches to ethics and differing understandings of what one's professional responsibilities can be seen to entail.

Our approach to ethics is grounded in feminist theories of care [24, 40, 100], the phenomenological approach to ethics [103] and a growing body of work that challenges dichotomies between emotional and rational ethical reasoning as developed with the concept of felt ethics [39]. We presume the multiplicity of ethical frameworks and the existence of competing ethical agendas. We assume that technology is not neutral but situated in political and social contexts [108]. Values are exercised in all parts of design and deployment of technology: from interaction within a design team to the deployment of technology and its application by users.

3.2 Data Collection

The empirical and conceptual findings of our paper were developed based on two distinctive sets of interviews with interaction designers and developers of AI[2]. Some of our participants are based in research institutions, while others work in the public sector, with a few working in private businesses or switching between academia, governmental agencies, and private companies. More details about the interviewees can be found in Table 1. All participants, except one HCI practitioner based in North America (the quote from this interview is marked in the footnote), worked in Scandinavian countries at the time of the interviews. The initial participants among interaction design practitioners were recruited from the authors' social networks. The initial participants among AI practitioners were selected through purposive sampling, recruiting among those who worked in the Swedish public sector and had at least ten years of experience. The rest of the participants in both data sets were recruited through snowball sampling.

All participants were given information about the study and an opportunity to ask questions, before signing an informed consent form. We emphasized that participants could skip any questions they preferred not to answer, and that they were free to withdraw their participation at any time. As the present research does not fall within the purview of ethical review in the country where the authors work, we have attended to ethical research practice through ongoing conversation among the author team and our broader community. In preparing the findings for presentation, we have either chosen not to include or have gotten additional approval for including details that

---

[2]We do not consider interaction design as separated from AI in terms of the technologies practitioners work on/with: some of our Interaction Design participants used AI-based technology in their projects. Yet, we see these as two separate communities with different vocabularies and ways of reasoning.

Table 1. Summary of interview participant demographics.

Gender	Female **(9)**, Male **(14)**, Non-binary **(0).**
Background/education	Engineering **(14)**, design **(5)**, economy/business **(2)**, other **(2).**
Job title/Role	Academia **(12)**: Associate professor (2), postdoctoral researcher (3), PhD student (3), research engineer (4).
	Public sector **(9)**: senior researcher (3), product owner (1), senior advisor (1), IT strategist (1), research director (1), project leader (1).
	Private company **(3)**: product owner (1), research director (1), developer (1).
Years of experience	1-5 years **(1)**, 5-10 years **(11)**, 10-20 years **(3)**, > 20 years **(8).**

are sensitive or might disclose an interviewee's identity or the organization they work for. Out of concern for anonymity, the description of the interviewees' professional roles is not uniform across the text, yet sufficient to understand their positions in the respective organizational hierarchies.

3.2.1   Data set 1. Interviews with interaction designers based in academia. The first data set consists of 10 semi-structured interviews that the first author conducted with academic interaction design researchers. Details on the interviewees' backgrounds are provided in Table 1. The interviews lasted one hour on average. They were centred around details of the participants' design work and the aspects of their work that they perceived as ethically challenging. The interview protocol included open-ended questions such as "Have you ever encountered any ethical issues in your projects" and "Have you encountered any ethical issues during planning/implementing/presenting your project". Major parts of each interview were centred around follow-up questions prompted by participants' initial answers, engaging with the details that participants viewed as ethically challenging.

3.2.2   Data set 2. Interviews with AI practitioners in the public sector. The second author conducted 10 semi-structured interviews with AI practitioners. The interviewees mostly worked with developing AI systems for government agencies. The agencies were responsible, among others, for the distribution of social benefits, taxation, and management of natural resources. The interviews lasted 40 minutes on average. The data was collected as part of a research project on the enactment of ethical principles by AI practitioners working in public organizations. The interviews began with open-ended questions, such as "Are you trying to make AI systems 'ethical'? If so, how do you do it?", "What do you understand by 'ethical AI'?" and "What does it involve in practice to develop 'ethical AI'?". Afterwards, the EU *Ethics Guidelines for Trustworthy AI* [1] were used as a prompt to engage with the interviewees' perceptions of ethical values in everyday practice. Examples of questions asked in this second part of the interview include "Did you establish mechanisms to ensure fairness/transparency/accountability in your AI systems?" or "Did you assess whether there could be persons or groups who might be disproportionately affected by the negative implications of developing/deploying such systems?"

3.2.3   Complementary data collection. The interviews were initially conducted as two separate projects. Our motivation to combine the data-sets stemmed from the significant overlaps in how participants approached responsibility for taking ethical action. After a preliminary round of analysis of differences and overlaps between the two data sets, we conducted three additional interviews with AI practitioners outside of government agencies and four follow-up interviews with interviewees of Data set 2. The follow-up interviews were aimed at targeting the interviewees' personal attitudes towards ethics. The interviewees were asked for examples of work situations where they could not act in line with their views (in situations of *ethical distress* [17]) and prompted to describe how they behaved and reasoned in such situations.

3.3 Analytical Process

The first and second authors had initially analyzed the data separately through thematic analysis [14]. After discussing the findings we had generated in this way, we combined the two data sets and performed a new, collaborative thematic analysis [20], encompassing the research materials as a whole. First, we analyzed three interviews together to agree on how to approach coding the materials. After that, we coded two interviews independently and synchronized the codes. The rest of the data was coded separately, with regular discussions about the themes we were generating and further directions for analysis. We performed inductive back-and-forth analysis to refine the codes, establish common patterns, and probe our hypotheses. Eventually, in collaboration with the third and fourth authors, we centred our analysis around different discourses underpinning the interviewees' reasoning around attributing ethical responsibility. The three ethical stances described in Section 4, as well as the elaboration of their connections, the emotion work they entail, and the work that they perform for practitioners in low- and high- scale environments, were generated from the interview data as a result of this thematic analysis process. While our research materials are not suited for advancing a systematic comparison between practitioners working in different domains, in presenting the findings, we indicate which data set each quote is drawn from so as not to obscure emerging distinctions between accounts from different types of practitioners.

3.4 Limitations

We acknowledge that our research materials have certain limitations, some inherent to all interview research. First, we cannot, nor are we trying to, provide a comparison between different 'groups' because (1) participants were not selected to represent a particular community or organizational setting and (2) the questions asked in the two sets of interviews differed, even though we intentionally established similarities through complementary data collection. Second, there are important limitations to what can be established from interview accounts without observing participants' activities *in situ* or studying the organizational practices and structures that shape their work settings. While a more extensive organizational study would be worthwhile, it is beyond the scope of this paper. Limiting ourselves to the analysis of interview data, we focus on practitioners' accounts of ethical responsibility rather than their practices.

4 WHO SHOULD ACT? THREE ETHICAL STANCES

We now turn to the three ethical stances we synthesised through our analysis: The *I-stance* is an individual approach focused on personal action. The *we-stance* attributes action to the collective, leaving the role of an individual unclear. The *they-stance* outsources ethics, positioning ethics as someone's else responsibility. The stances should be understood as discourses that underpin answers to the question of *who should act* rather than as individual dispositions. Taking one of the discursive stances is not bounded to an individual: referring to different situations, one person can shift between all three stances. The stances are related to but not solely defined by the use of I, we- or they- pronouns. Most importantly, they differ in emphasising different actors as the subject of action. In this section, we present each stance as an analytically distinct discourse. In section 5, we will discuss the relationships between the three stances, the emotion work that comes with them, and how they play out in different contexts.

4.1 I-stance

The I-stance presumes that a person is an agent of ethical action. Therefore, it attributes high value to personal decisions. This stance places the least possible distance between an individual's actions and their consequences. It implies a belief that a person can influence relevant outcomes

and, hence, can (and should) take the blame or distress related to any potential negative outcomes – or the credit for success. In this section, we consider two types of ethical concerns in turn: first, matters of interpersonal interaction in technology development processes, and second, issues that have to do with the technologies being developed. While these two ethical domains are different, we highlight that they intertwine in the development of technology since technology ultimately changes relationships between people, even if issues of interpersonal interaction may appear distant in settings complicated by organizational hierarchies. We start presenting the I-stance in the domain of interpersonal interaction where it is represented the best. However, the I-stance can be taken in any domain and at different scales: from interacting with colleagues and research participants ethically to striving to develop a non-biased AI. While practitioners in different settings are faced with different ethical questions, the common denominator here is taking personal responsibility.

4.1.1   I-stance and interpersonal relationships. The I-stance is familiar to researchers as it often needs to be taken when building relationships with research participants [95], creating safe design spaces [4, 11, 25, 82, 98], and steering research projects in a desired direction. These are all settings where individual ethical decisions have a high impact and, thus, the I-stance is at its most powerful. This is illustrated in the following interview excerpt from an interaction design researcher reflecting on the importance of daily work with co-designers of technology. The interviewee highlights the importance of her personal role in setting up trusting relationships with project participants. She stresses that her personal role matters despite the existence of ethical guidelines that outline the direction of action, since guidelines alone are not enough for implementing a project in an ethical way or building interpersonal relationships within a team:

> "There are national ethical reviews, guidelines, but it's more like the daily practice of how you're doing research. I have examples of ethical ways of working with myself, my own body. But in this case [the project discussed in the interview] it was how do you engage other people in your work and create a space where everyone feels comfortable about sharing their opinions through words, or through movements, you know, like a place where you can create trust." (Interaction design practitioner, early-career academic, female)

The same interviewee talked about the importance of daily work practices and the ethics of everyday decisions within the design process, such as communicating with teenage research participants. By law, teenagers are not yet fully fledged subjects and do not have to be asked to consent to participate in the study. Instead, their parents should consent on their behalf. Yet, our interviewee chose to communicate with the participants directly, too. What guided their decision to ask the teenagers themselves was a feeling and intuition rather than a formal rule:

> "They say: it's their parents, who need to sign it for the minors: if you're under 18, um, then it's the parents who sign. Like, it doesn't matter what the minor thinks, the parent needs to sign it. So I had two documents [informed consents], because I thought that was a bit weird, you know, because these were teenagers, they were almost 18." (Interaction design practitioner, early-career academic, female)

In addition to this instance of a researcher going beyond formal ethical requirements to act in a manner they considered ethical, our data contained multiple further examples when the I-stance was applied in research ethics in ways that the participants deemed successful. However, the personalised take on ethics became demanding and problematic for an individual when touching on structural issues or power imbalances. In such situations, the cost for the individual taking action becomes too high, and operating from an I-stance can have high emotional costs without leading to action.

Our second example demonstrates an extreme case of a personal crisis that resulted from a situation of ethical distress. The interviewee felt obliged to act but, in the end, did not act because of the high personal cost that taking action would have entailed. In the quote, the participant refers to an internship at a company where they spent parts of their time as a doctoral student. The interviewee strongly disapproved of the company's work practices. This made working for the company an ethical problem and left the participant balancing the intention to act with the fear of negative consequences:

> "Should I say something to anyone? And then I realised that the guy who has a company, he is pretty powerful figure [...] and he has a lot of connections, and I didn't want to start moving things there. I mean, it wasn't my job in that case. I stayed there for three months and my whole self was destroyed, because I returned back and I started feeling that I should have said something and I didn't." (HCI practitioner, early-career academic, female)

Experiencing emotional struggle, in their case, led not to an action but to a personal crisis related to the feelings of losing one's agency. This kind of strong emotional response – although typically in a less damaging form – was common for interviewees who took personal responsibility for the consequences of their actions while finding themselves in situations where they perceived acting to have too high personal costs.

4.1.2   I-stance in the development of complex technology. Operating from the I-stance can be challenging when multiple parties are involved. In our data, this was especially relevant in accounts related to AI development. AI systems exist in complex environments compared to the contexts where designers and users meet face-to-face. The complexity stems from the fact that many participants are involved in the development of AI, from data labelling and developing mathematical models to applying AI-based technology in practical settings. Nonetheless, the process of AI development, too, features situations where individuals operate from the I-stance.

The following example describes a situation when an interviewee, who occupies a senior position in his organization, was planning to install a facial recognition system in a shared, open-plan office. The interviewee, who led the project, explained that they felt uncomfortable thinking about how his colleagues would react to this facial recognition technology. He explained that his main reason for *emotional discomfort* was the realisation that computer scientists are sensitive to privacy issues, and it would be difficult to communicate how the system would not store any data. It is interesting how the interviewee makes a distinction between the 'ethical' and the 'emotional'. To define the problem as ethical, they refer to their feeling and other people's feelings rather than a rule, but this makes them initially doubt whether the issue is, indeed, ethical:

> "I don't know... ethical? It was just also just... emotional. I mean, how, how do people feel about having some system taking their picture and processing it and, and how do you communicate that: no, it's okay because we don't store the pictures. (...) it has the nature of some kind of surveillance, right? No, but it feels like it's: so you have a camera and whenever I pass by that, you will record that. (...) That particular demo machine was just outside my office, and I wanted it to be in play. And there was definitely an ethical problem to force that upon [others], because [...] There are slightly more people within computer science who are more sensitive about such issues, who are very careful about computer security and integrity and all of that." (AI practitioner, senior academic, male)

Another example comes from a researcher who worked on developing machine learning algorithms in a commercial company. Given that the company had other specialisations than developing AI,

the interviewee's team had to explain their work to other teams. Reporting the results of ML models can sometimes be challenging [104] because the statistical results of the model evaluation can be easily misinterpreted. The interviewee found himself in a situation where the results reported by their team were misinterpreted by others (the 90% accuracy rate model was treated as a sign of success, which it actually was not). At that time, the team was facing criticism from management, so the presentation was a high-stakes event. The interviewee described feeling conflicted because he wanted neither to mislead colleagues nor to ruin relationships within the team:

> "If we know that people can misunderstand that 90%, we're responsible to make sure that, you know, they don't hype it too much or disregarded too. We are responsible for that. So even if the team is in a tough spot and people are putting pressure on it, I don't think it's like really ethical to go and say but we got 90%. (...) I wasn't sure, but in the end I was like 'okay let's do this' [say that 90% was not a success], but it was risky. I actually went and spoke with the team, but it didn't go well for me. Exactly what I expected happened that people were thinking 'this person wants to act all like ethical and sabotage the team' and it was kind of dismissed." (AI practitioner, research engineer in a private company, male)

Taking the I-stance and accepting personal responsibility can be problematic when action cannot be taken or the costs of taking action seem too high. What is more important, the outcomes of taking the I-stance in situations of developing complex technology were not necessarily beneficial when individuals did not have power to act upon their will.

4.2 We-stance

The we-stance considers a collective – an organization, a company, a governmental agency – as a unit of ethical action. The individual is considered a part of this collective subject rather than an independent actor. When technology is produced in large organizations with a strict distribution of labor, taking an I-stance towards ethical action may not be feasible. In such cases, taking the we-stance – where the organization as a whole is seen as the unit of action that is responsible for acting ethically – may seem like the default choice for the individual practitioner. The we-stance corresponds to a sense of belonging to a group, as illustrated by the next interviewee who shares the collective responsibility of sustaining the reputation of civil servants:

> "In [the country] we have a very strong type of culture within the administration. For many, many years people have had a huge trust to it. It has been open and we have built that in many hundred years I would say. So, we need to maintain that within the digital age." (AI practitioner, senior advisor in governmental sector, male)

The we-stance is not specific to government officials and AI developers. Interaction designers in industry settings have also been documented to frame their responsibility vis-à-vis the organization in this way, as demonstrated by Lindberg and colleagues [64, p. 5]. The main distinction between the two stances is that the role of individual responsibility is not emphasized in the we-stance, and not positioned in conflict with a collective. Individuals taking the we-stance consider themselves part of a collective but do not necessarily see a direct connection between themselves and any necessary ethical action. In the interviews, the collective take on ethics could often be noted through the use of the personal pronoun 'we' instead of 'I', along with expressions referring to the group, such as 'in our team/organization' or 'as a company'. The use of the pronouns 'I' or 'We' is not the only indication we rely on in categorizing interview accounts as instances of a particular stance. Rather, in recognizing the we-stance, the emphasis is on the absence of any specification of an individual's personal role. In contrast, in identifying the I-stance, the key is whether interviewees describe

themselves as responsible. The distinction became especially noticeable when interviewees who were asked directly about their individual opinions defaulted to collective 'we' answers:

> "-In case you would experience a tension between, for example, you and your projects or your organization, what do you think you would rely on for solving that problem? For example, would you rely on your inner feeling or go to maybe some documents and guidelines and procedures or talk with other colleagues or with experts?
>
> - So, how would **I** go about fixing that? That's something **we** do pretty much every day when we run a company [...]." (AI practitioner, research director in a company, male)

Referring to a company as the actor holding responsibility in front of another collective — the 'customer base', users, or stakeholders – is one example of operating from the we-stance. This is also a common discourse in *AI Ethics*, which often considers groups and organizations as the relevant units of action [88]. These two quotes, from different participants, provide further examples:

> "I think we also might have an educational obligation as a company. We should actually inform, in some sense, our customer base what it means to buy a data-driven product." (AI practitioner, research director in a private company, male)

> "[Interviewer]: So, how is transparency applied for example in the systems or in the methods?
> [Respondent]: Again and I would say that's by education because it's like a two front war if you want. Because we have to educate our clients and stakeholders at the same time as we are trying to make systems for them." (AI practitioner, research director in a private company, male)

The we-stance does not emphasize an individual's role in taking action, so it does not provoke internal conflicts between personal intentions and external circumstances. In that sense, to the individual, operating from the we-stance is a safer choice, as it can help them to avoid both the risks related to taking action (losing one's job or ruining workplace relationships) and the emotional burden of non-action. Because the we-stance distances the individual from ethical action, interview accounts that use the we-stance are less emotionally charged than those given from the I-stance. Most of the examples of ethically challenging situations discussed within the we-stance include, for instance, educating users – a significantly less emotionally challenging activity than the situations described with the I-stance. User education is considered a step towards developing ethical AI because it increases users' awareness of potential issues related to data-driven products. This educational approach transfers some part of ethical responsibility to the product's users, implying that they should be aware of potential negative consequences resulting from its use and, further, should stop using products that lead to excessive negative outcomes. Overall, when interviewees speak from the we-stance, their accounts involve less emotional expression in words and intonations, both in what is said and *how* it is said. The lack of emotion may be specific to our data set, where collectives were mostly formal workplace collectives or units of bureaucratic organisations rather than self-organised groups.

In our data, strong emotional responses were noted when the participant personally witnessed negative consequences of technology. In the following case, which we consider transitional between the I- and we-stances, the interviewee was working in a team that used a machine learning algorithm for facial recognition – originally developed for other purposes – to build their technology. As it turned out, the algorithm, developed by a third party, did not recognize participants with dark skin during the testing stage. The bias was discovered at the latest stages of the process when the team finished the artifact and started to use it as part of their work. Despite the discovery, it was decided to continue the project – a decision that resulted in the interviewee's feeling of moral discomfort.

> "The device was finished, the vehicles finished, it took months to create and someday you have this [discovering the bias] we cannot start the project over again. The discussion was like 'we have to be careful (...) One month ago we were doing another thing with [...] and two students came to see how they work, because I was going to give them the cameras and everything. I didn't remember that one of them was black. I forgot about that. I was showing them how to use it. [...] And then he's like 'can I try, can I try it?' And it doesn't work. And I was like 'Oh, I don't know.' We were very ashamed." (academic researcher)

In this case, the problem of bias was not abstract but related to personal face-to-face encounters with users from an excluded group. The interviewee described the feeling of shame they experienced when showing the device to the user. The way the participant talked about bias also visibly expressed discomfort. However, they do not separate themselves from the team as a collective subject and do not attribute the responsibility for developing biased technology to themselves personally. They talk about shame as a shared emotion and attribute the responsibility for the problem to the team as a whole, who attempted to remedy the situation by contacting the algorithm developer, but did not discontinue the project because the cost – in terms of money and invested resources – appeared too high.

In general, the we-stance puts fewer demands on personal judgement compared to the I-stance. In the realms of formal institutions, it considers ethical decision-making a result of following the established set of organizational procedures rather than the ethical judgements of a reflective practitioner. In commercial and bureaucratic organizations, it implies the need to translate ethics into formalised language of organizations: into guidelines, norms, procedures, and metrics. When there are no guidelines at play, and ethics 'is not measured,' ethical issues become intangible – as noted both by other researchers [64] and our interviewees:

> "In all of this, to also keep track of something that is quite elusive as ethics, I think you need to actually bring it down to a measurement so it's easy to see. Preferably in a dashboard or some spreadsheet or whatever that people could look at. Perhaps this might sound crude, but put a price tag on it." (AI practitioner, research director in a private company, male)

4.3 They-stance

The last ethical stance we discuss is the they-stance. The they-stance externalises responsibility for taking ethical action. Here, ethics was not claimed to be unimportant, but participants acknowledged it as *someone else's* responsibility, be it a particular group of developers, ethicists or social scientists, users of the resulting system, or another group that the technology practitioners did not belong to. Below, we report the most common rhetorical techniques that served to distance our participants from ethical responsibility (whether as an intentional move to avoid responsibility or not).

As a first technique, participants distinguished between the 'ethical' and the 'technical.' The quote below refers to a work domain where the end-users are other engineers. The interviewee draws on a distinction between 'ethical', which has to do with people, and 'technical,' which has to do with interaction between machines and, thus, is not taken to imply any ethical considerations. This removes the possibility of recognizing an ethical concern in 'purely' engineering matters:

> "Many of the applications we address within [our organization], they have to do with looking at images of cracks in composite plastics and assessing them in different ways. There aren't that many ethical concerns in that. Or maybe if you help me, I might find [the concerns], but often not. Many applications are very technical, or they have to do with process automation, industrial process automation, take the readings of

various sensors and do something with these. These ethical concerns, typically, of course, come when you have to do with people in one way or another." (AI practitioner, senior researcher in governmental sector, male)

Another technique implies limiting the domain of ethics to end-user interaction. Such considerations were expressed by an interviewee who developed an AI-based system for government agencies. Here, the developed system is not solely a machine-to-machine application – and it will be used by other people who are not engineers – but it is not explicitly designed for the end-user (i.e., citizens). As expressed in the quote, the presence of an intermediary – an administration – implies that ethical considerations should be the concern of the client, not the developer:

> "I mean our work in this project is more about the core technologies and we also support the government agencies when they build their in-house applications. But we don't take responsibility for the applications that are directed out to the citizens. That's up to the government agencies, and that's their responsibility and their ownership of those techniques." (AI practitioner, senior researcher in governmental sector, male)

This logic could be applied to any service that does not directly interact with end-users. The interviewee in the following quote applies a similar technique when he talks about a service developed for the needs of the governmental agency, claiming that the service is not subject to considerations regarding transparency because it is a time-saving intermediary device. Yet again, this discourse implies that the tasks of 'mere optimisation' are free of ethical considerations – a type of reasoning that imagines technology as neutral and that has long been challenged by *Social Studies of Science* [12, 41, 47, 108]. While ethical considerations may arise, the responsibility for identifying and addressing them is assigned to a third party who interacts with end-users. This leads to overlooking the potential responsibility of both the system itself and, more importantly for our discussion, the developers of the system:

> "So they do email classifications, so when people ask questions by email they try to classify which topic the email is about and then they route that to the correct administrative person. And I don't really think that transparency enters into that process at all because it's just a sort of a filtering step. We don't actually make any automatic decisions based on the system. It's only like routing the messages to the correct person, so it simply just saves time for the administrative part of [a government agency]." (AI practitioner, senior researcher in governmental sector, male)

Finally, our last example represents a slightly different technique where an ethical problem is recognized, but it is attributed to a particular domain unrelated to the area of the participant's job. Working in certain domains is acknowledged to be ethically problematic (such as recommender systems or automation of work), but the position is taken that ethical concerns can be avoided simply by not engaging with these sensitive AI applications:

> "- have you ever encountered any ethical issues, during any of these projects or any other project?
>
> - Ethically? Like how, for example? I mean, if I was doing something like for automation factories or like self-driving cars maybe I would have … I don't think I will do it. It depends on the… cause like when you work in this kind of stuff, there is always like the question 'who are you working for?'. Especially in AI. I have friends that went to do internships at Amazon. They don't have a problem with that, but I will not do it." (AI practitioner in a research institution, male)

Within this approach, ethics is recognized as relevant only to particular domains, specific AI techniques, or applications thereof. Therefore, direct involvement or action is considered optional by those who do not work within those realms.

5 EMOTION WORK AND THE THREE ETHICAL STANCES

In the previous section, we considered three ethical stances – discourses that attribute responsibility for ethical action to an individual, an organizational entity, or *someone else*. The stances are not fixed individual dispositions: the same person may shift between different stances even within a single conversation as they produce an account of their actions and experiences. While the stances intertwine, they are analytically distinguishable and perform different work in attributing responsibility to specific actors. The most crucial distinction is between the I-stance, which places ethical responsibility on an individual, nudging them to act, and the they-stance, which help individuals avoid feeling obligated to act through techniques that transfer responsibility to another actor.

With the help of the concepts of *distancing* and *vulnerability*, we now consider the relations between the three stances and the emotion work that operating from each can place on technology practitioners. We also analyze the work that the stances perform for technology practitioners operating in different settings, making a distinction between *low-* and *high-scale* environments.

5.1 Vulnerability and distancing

Vulnerability in HCI is commonly considered to be a characteristic of a group affected by technology or social structures, but, here, we apply the concept to technology practitioners who may end up in vulnerable positions while (and because of) trying to act ethically. By vulnerability, we refer to the risks related to direct personal engagement in the outcomes of one's action [82]. Being vulnerable, then, implies carrying on emotion work [4, 49] as a part of technology design and development. Vulnerability implies both emotional burdens and material consequences, such as the risk of losing one's job or entering into a conflict with a powerful figure, an action that can lead to multiple negative consequences.

The I-stance is prone to evoke vulnerability in situations of asymmetrical power relationships. As we illustrated in the Section 4.1, taking action in such situations can mean, for example, risking one's job, whereas not acting can cause emotional distress. The interviewees who had to act against their ethical views reported feelings of losing one's agency, accompanied by emotional discomfort of varying intensity, including feelings of guilt, sadness, and shame (as illustrated in the quote in section 4.1.2 where an academic researcher expresses having felt really ashamed). In the most extreme case, the same interviewee who felt that her 'whole self was destroyed' (see 4.1.1) due to not taking action, discussed feeling guilty because of her role in enabling the status quo:

> "They [the company] will manage to get more funding only because I was there. Of course, if I wasn't there, they would have done it themselves, it wasn't something… it was a manual work. But still I helped. […] I felt trapped and I could, I explain that I felt trapped. I couldn't speak, I was scared and so on. […] And at the moment I was like, I was crying." (HCI practitioner, early-career academic, female).

Another example of feeling sad and frustrated comes from a researcher who, because of institutional pressure, had to step away from a research project she wanted to engage in:

> "You know, I was really sad that I was told to not do it because I was really excited about it and I felt, you know, this was a research question driving me with curiosity, and I was told to not do it. So, I was like, okay, what should I do then? And then I felt

like, then I did turn on quite a different track to do something else. You can definitely imagine frustration" (Interaction design practitioner, early-career academic, female).

In these cases, the interviewees were economically and emotionally vulnerable: while recognizing ethical concerns, they also recognized that they would face negative consequences for both acting and not acting – either in the form of losing their employment or in terms of experiencing the loss of self. Ethical responsibility in such cases was not a question of distributing duties within an organization, but something that was *felt* – sometimes intensely – as a somatic experience of discomfort.

In contrast, the we- and they-stances evoke a significant degree of emotional *distancing*, which is achieved by situating subjectivity outside of the individual. The we-stance considers the collective – an organization, a company, or a community of AI developers – to be the relevant subject of ethical action without specifying a personal role for the individual practitioner. Within the we-stance, distancing can be additionally achieved through rules and formal procedures that reduce the need for personal judgement to a minimum. These distancing mechanisms move ethical responsibility away from an individual by turning ethical decision-making from a felt phenomenon to an organizational one.

Within the they-stance, distancing is achieved by externalizing ethics to a third party. The techniques that we encountered in the interviews included limiting the sphere of the 'ethical': setting up a dichotomy between 'ethical' and 'technical' issues, excluding everything that does not have to do with 'people' (end users) as well as excluding service intermediary technologies from the domain of ethical concern, or assuming that only specific technologies belong to 'ethically sensitive' domains. Following this logic, the need to consider ethics is acknowledged only in certain situations – and commonly, these do not include the domains of the interviewee's employment. Within such a discourse, the need to take ethical action is either not recognized at all or it is attributed to a third party. In our interviews, this third party was usually a client who interacted directly with end-users, but it could also be the end-users themselves, a person responsible for ethics in the company, or a particular algorithm [89]. The ultimate example of distancing is what Seaver has labelled as 'decorrelational ethics' [89], a move to fulfill the dream of creating a technology that avoids value conflict altogether, thus eliminating the need for conversations about ethics.

5.2 Low and high-scale technology design and production

The stances perform different work in *low-* and *high-scale* situations of technology design, where they can either help practitioners see the connections between their actions and outcomes or become discouraging by placing unrealistic expectations on an individual who lacks power to act upon them. By low-scale, we refer to situations with a limited number of participants, where practitioners have more control and power over the development of technology. Following Brown and colleagues, we presume that these situations usually occur in the context of vertically integrated designs [15] that do not depend on broad ecologies of technologies that are interdependent (such as the Google suite of services). Instead, they are intended for a known context, where practitioners can control the use of the system, and they are meant to be used only by a limited number of people. These are often the kinds of situations where practitioners can directly interact with the intended users of the technology.

Unsurprisingly, the I-stance seems to fit best in situations of low-scale design work, where an individual has power to act according to their values. In our interviews, taking a proactive I-stance in controlled circumstances – for example, when interacting with research participants who were also the intended users of the technology in question – allowed practitioners to achieve desired outcomes and experience a sense of fulfilment from work well done: establishing a trusting atmosphere

withing a project, treating co-design participants with care and respecting their autonomy (see the accounts from an interaction design practitioner in 4.1.1 and 4.1.1). Such interactions were often related to a context where the consequences of one's activity were directly visible and the researcher had sufficient resources to act. Yet the successful examples of individual interventions were not common beyond the scope of small research projects. Despite good intentions to serve communities, technology practitioners were often discouraged by their lack of capabilities to induce change. The following quote illustrates the shock this can entail:

> "I did my undergrad in electrical engineering, where I learned how to do machine learning, how to design circuitry and devices, technology so that they work faster. So all of these are very, very important problems to solve from a science and engineering perspective, but from where I am coming from – small town, surrounded by very rural villages – their problems were not addressed in my education. So it didn't reflect. Like I was learning, I was really working hard on my education, but that wasn't like, that didn't feel very directly useful for the community I belong to... [...] that was really a shock for me! I was in my first year, I was so upset that my politics failed." (HCI practitioner, early-stage academic, female[3]).

In contrast, the they-stance was usually taken when the implications of the technology could not be directly observed or felt or when the practitioners clearly lacked resources to address the issue. When the implications became visible for the practitioner, such as in the cases of observing the effects of a racially biased algorithm (see 4.2) or setting up a system that could threaten colleagues' sense of privacy (see page 10), interviewees often switched to describe situations from the I-stance. Shilton has reported similar findings from her fieldwork, noting that software engineers could more easily relate to privacy and other ethical issues after testing the system that they were designing [91].

In situations of high-scale technology development, actors are separated from each other by the power hierarchy within an organization, and most of them do not have much control over the course of technology development. Any case of developing AI/ML-based technology would be a high-scale technology development, given the amount of participants involved at each stage of the project, from data collection, labelling and storage, to algorithm deployment. The I-stance becomes problematic in such environments, since the high personal costs of operating from it do not necessarily pay off in terms of meaningful impact. Accepting responsibility will not necessarily lead to action if multiple other actors are involved, the action feels too risky, or mechanisms for action are unclear. The overly individualistic approach to *owning ethics* [66] has a clear risk of becoming too demanding, overwhelming, and ultimately not actionable if the sense of responsibility is not paired with possibilities to make a change. Given the difficulty of implementing individual values in relation to complex AI technologies, operating from distanced we- and they-stances is a safer and more reasonable (from the individual point of view) choice for developers employed by large organizations. Distance creates safety by removing the uncomfortable and potentially paralyzing burden of inescapable uncertainty ('Am I acting ethically?').

Further, they- and we-stances often translate ethics into an organizational phenomenon: a matter of guidelines, rules, and bureaucratic matters. While we certainly do not advocate for relying solely on emotion in managing ethical issues, the opposite, too, can be problematic. When ethics is approached *only* through organizational lenses – quantified and measured, not felt – it becomes detached from lived experience and it can no longer be experienced as a field where personal action matters. This alone is problematic as it can lead to feelings of losing one's agency and sense of control. Distanced stances reduce feelings of personal involvement and meaningfulness, creating

---

[3]The participant was working in North America during the interview.

detachment that can be damaging for a person [49]. Considering ethics solely through organizational lenses, then, overshadows the lived experience of people affected by technology, along with those who are developing algorithmic systems. Given the contextual nature of ethical action that cannot be adequately formalized through a system of rules [40, 100, 103], this is problematic for efforts to develop technology ethically. Strict reliance on organizational understandings of ethics poses the risk of ignoring broader questions, such as asking whether the technology should be built in the first place [97], since considering such questions is unlikely to be part of organizational guidelines and, thus, a matter of no one's responsibility.

6   DISCUSSION: BETWEEN VULNERABILITY AND DISTANCING

Our analysis follows the tradition of studying ethics in practice: approaching ethics on the technology production site as a matter of concrete technical decisions [91], situated in organizational contexts [13, 42, 64]. With a few exceptions [31, 64, 95], prior studies have focused on the work of industry practitioners based in the US. We broaden the state of the art by interviewing technology practitioners who are located in Scandinavia and by including participants who work in the public sector and in academia. Crucially, our research furthers the exploration of the emotional aspect of ethical responsibility, started by researchers in Interaction Design [4, 50, 82] and scholars of Ethics in Practice who have analyzed how recognizing ethical concerns affects technology practitioners' feelings [99, 107]. We contribute to the study of ethics in practice by focusing on the emotion work related to recognizing an ethical issue and accepting (or refusing) the responsibility to act on it. While the importance of developing ethical sensitivity – the ability to *recognize* ethical issues – has been studied and documented thoroughly in prior work [13, 64, 91], we focus on the next step: taking responsibility for ethical action. We argue that the emotional aspect of attributing responsibility is central to practitioners' subsequent choices to act or to refrain from taking action.

6.1   Taking ethical responsibility involves bearing an emotional burden

Our findings describe how the three stances place a different degree of responsibility on the individual practitioner, either liberating them from any personal need to act or putting them on the spot. While taking the actionable I-stance may seem empowering in that it positions the individual with high responsibility and high personal involvement, it also makes the individual vulnerable. Taking more distanced we- and they-stances can, in contrast, help individuals to avoid emotional discomfort, at least in the short term. With the we-stance, the interviewees were able to speak about 'we' or 'us', defining themselves as part of a collective subject while avoiding a direct personal need to act. Within the distanced they-stance, we have noted three separate *distancing techniques*, which our interviewees deployed to frame ethics as a matter of someone else's responsibility: (1) distinguishing between the ethical and the technical; (2) limiting the domain of ethics to interactions with end-users; and (3) marking only certain areas of technology application as ethically problematic and, thus, as requiring ethical reflection. These distancing techniques also explain why participants come to *not* feel the need to carry ethical responsibility in their workplaces.

We highlight that the very fact of noticing an ethical issue and acknowledging one's personal role in acting puts emotional burden on the individual. Due to the high emotional costs of responsibility, it is reasonable that accounts from the I-stance were not that common among our participants. Here, our findings align with the research of Widder and colleagues [107] that reports on emotional distresses and sensations of anxiety, depression, and isolation related to raising ethical concerns in tech companies by workers who had no actual power to address such issues. Similarly, when taking the I-stance, most of our interviewees ended up feeling distressed: sad, ashamed, frustrated, or left with a feeling of 'losing oneself.' In our data, positive examples of taking ethical action came

from the area of research ethics, especially from interactions with participants in participatory design and co-design projects where the interviewees had the power and resources to act in line with their ethical sensitivities and to work on establishing trusting and respectful relationships.

Taking the I-stance could certainly be beneficial for the development of more ethical technology, yet given the high cost of acting upon ethical issues, it is only reasonable that our interviewees deployed a variety of distancing techniques. Given the similarities between our findings and the findings from Widder and colleagues [107], we can say that individuals located in different institutional and regional settings – yet in structurally similar circumstances where they lacked the power to address their ethical concerns – were experiencing similar feelings of distress, anxiety, or disappointment. This structural shaping of how emotions come to matter for ethical (in)action is a central part of why emotions should not be dismissed from conversations about ethics as individual and volatile matters.

Both we- and they-stances allow interviewees to distance themselves from ethical responsibility, hence largely avoiding negative emotional consequences related to even recognizing an ethical issue. The distanced stances relieve practitioners from discomforting feelings of powerlessness, responsibility, and internal conflict related to the discovery of injustice. But there is a danger in both taking the I-stance and not taking it. The danger of being too vulnerable relates to the possibility of experiencing an emotional crisis and/or facing material consequences for acting. The danger of distancing, on a psychological level, manifests in the form of alienation and detachment – something that Hochschild noted in workers who had to engage in emotion work in the service industry [49]. At scale, the danger of distancing lies in that it may enable broader negative outcomes, such as the production of harmful technology.

6.2 Steps toward making ethics actionable

While we make an analytic distinction between situations of low- and high-scale technology development, there is no opposition between the scales: all high-scale projects demand personal involvement. Large AI systems are built by workers who pay attention to seemingly 'small' things, such as cleaning data sets [80, 89]; ethical guidelines are driven by enthusiasts fighting their way through organizational structures [65, 84]; the implementation of otherwise abstract principles is made possible by very concrete people [78, 109]. The I-stance has to be taken in order to make high-scale settings work, but taking the I-stance is likely to have serious negative consequences for individuals with limited power. That said, we join claims to refrain from framing ethical matters as matters of individual decision-making and responsibility [8, 87]. Our suggestions align with prior calls to develop more supporting structures for practitioners who take responsibility for acting ethically, both in low-scale projects and within big tech companies [65, 83]. In what follows, we discuss three strategies to make ethics actionable: normalizing *a degree* of vulnerability and discomfort, developing organizational structures that support individual ethical action, and developing hands-on approaches to demonstrate the limits of individual action at the technology production site.

6.2.1 Normalizing vulnerability and discomfort. Taking an actionable stance is needed to implement change. Therefore, our first conclusion is that we have to normalize the vulnerability of the I-stance along with a degree of discomfort that follows from recognizing ethical issues. Inspiration for becoming *more vulnerable* and shouldering more personal responsibility can be taken from first person perspectives in interaction design, such as soma design [52, 53], that strive to reduce the distance between the creator and their work, and encourage approaching one's inherent vulnerability not as a problem that needs to be hidden but as a ground for connection and a reason to care for each other [82]. Approaching design work from a vulnerable standpoint means accepting

the value of our felt experiences, discomfort, and failures [55] and respecting gut feelings – the not-yet-fully formalized *discomforts* and sensations of unease that often point at problems which were not yet articulated. Such an approach to social interaction on the technology production site implies building new grounds for collective action, based on ideas of the inherent relationality (rather than a supposed autonomy) of human beings. This aligns closely with ethics of care [24, 100].

Normalizing vulnerability means accepting that a degree of discomfort will always be part of an ethical practice of technology production, especially in high-scale settings. Further, even within low-scale projects, an individual's control is limited. For example, design materials carry their own histories [23] which can introduce traces of colonialism and oppression even into the purportedly most ethical and justice-oriented designs. Accepting and articulating the discomfort of not having total control over the process while still feeling called upon to bear responsibility for its outcomes can be the first necessary step in changing the status quo.

6.2.2 Developing structures to support ethical action. Besides normalizing a certain degree of vulnerability and the discomfort that comes with it, we need to develop structures to support ethical action. Otherwise, it is very likely that those who are in the least powerful positions will end up being the most vulnerable and bearing the biggest cost for taking responsibility. Vulnerability is not a meaningful answer when there is no safety and equality, or when urges to be vulnerable come from those in positions of power and security towards those who are located lower in a social hierarchy. Liboiron and colleagues provide an example of a situation where the request for being vulnerable evokes nothing but anger:

> "I was giving a talk about CLEAR's feminist science at a university reputed for its progressive politics. It hosted one of the most alienating Q&As I've ever been invited to (except for that time in that philosophy department). One question stood out. A woman asked me to share my failures in the lab. I asked what she meant. She said, "I want you to be more vulnerable." I obliged her. Later, I was furious. Furious for the question—asking me to be more vulnerable, standing at the front of the room as a small, Native woman whom academia and dominant science are built to erase, trying to work where there are few roadmaps. I was furious no allies in the room stepped in to intervene, despite their rhetoric of allyship. But mostly I was furious at myself for obliging instead of burning that house down." [62, p. 147].

Attempts to label ethical decision-making on the technology production site as solely an individual responsibility are dangerous for those who have little protection within organizational hierarchies. Vulnerability has to be equalized with the degree of power. When the outcomes of an action are beyond one's control or when one lacks the ability to make the necessary sacrifices, distancing becomes a matter of safety. While protocols, guidelines, and formal procedures are imperfect, at the lack of better means, they may serve to ensure the safety of individuals in structurally vulnerable positions within their organizations. However, solely relying on protocols is insufficient. It is crucial to have supporting structures that provide checks and balances, such as technology workers' associations.

6.2.3 Reckoning with the limits of individual agency. The conversation about ethics of technology is too often framed with an excessive focus on the individual as the main actor of change. This is in stark contrast to the distributed nature of technology development. To create a more realistic perspective on the value of individual efforts that are not supported by associations and other forms of political power, we suggest bringing forms of reasoning from *Science and Technology Studies* (STS) closer to tech practitioners, that is, making the consequences and limitations of individual action more readily *felt*. Attending to the histories of design goals and materials, including data,

can stimulate us to recognize the limits within which we can induce change and help us respect the unexpected consequences our work may have. We see an opportunity here to bring ideas that are almost taken for granted within STS, such as the agency of technological artifacts and politically motivated decisions that can underlie their design [108], closer to design and engineering practice.

In practice, this can mean developing *historical sensibility* [93] – demonstrating the historical, consecutive nature of technological development – through hands-on work with design practitioners. One successful example of this type of interdisciplinary translation is the *implosion* exercise that Dumit developed based on Haraway's scholarship [28]. In this exercise, participants are asked to pick up a simple artifact, like a pen, and answer a number of questions to trace its connection to the world outside of the classroom/workshop. The questions can be very concrete: *"How was it produced and who is involved in its production?", "What materials are involved in its production and maintenance? Where have these materials come from?"* [28]. An example of using the implosion exercise to uncover the connections between cultural discourses, media, and intellectual lineages in emotion detection can be found in recent work by Arnelid, Harrison and Johnson [3]. Another example is Joler and Crawford's [23] critical map of the Amazon Echo device which uncovers the supply chain of materials, planetary resources, data, and human labor behind the device, highlighting injustices and exploitation at each stage of the process.

What we suggest as a next step is experimenting with making the complexity and relationality of technology production as tangible and *felt* as possible to highlight the limitations of individual agency in relation to the complex histories of materials. We encourage CSCW scholars to work on developing hands-on ways to introduce historicism into the technology production site and to have conversations about ethics, not only analytically but also in a way that appeals to our felt sense as technology practitioners. This is not an easy task since high-scale technology by its nature produces distance between its users and creators and renders the underlying complexity of its production invisible. Yet, we believe such an intervention is needed to bridge the gap between our emotional understanding of ethics and the distributed nature of technology production. This can help us not only to avoid placing unrealistic hopes on the potential of an individual reflexive practitioner but also to recognize opportunities to make change collectively.

7 CONCLUSION

We have considered *emotions* related to recognizing ethical issues on the technology production site and attributing responsibility for acting on them. While the scholarly conversation about ethics has traditionally focused on searching for a rational ground for ethical action, we argue for the need to attend also to the emotional component of ethics as it helps explain why taking responsibility and even recognizing an ethical issue can be so problematic. Based on an interview study with technology practitioners from academia, technology industry, and the public sector, we have identified three *ethical stances* – I, we, and they – that technology practitioners use when providing accounts for taking or not taking ethical responsibility. Further, we have introduced the concepts of *vulnerability* and *distancing* to analyze the differences between these stances and the emotion work that they entail. Our analysis demonstrates that recognizing an ethical issue and taking an actionable I-stance towards it places a heavy burden on practitioners, especially when they do not have the power to act upon their ethical sensitivities. Cultivating ethical sensitivity, then, is important but not sufficient to enable productive ways of acting. We argue for the importance of building new organizational practices and safe conditions for taking personal responsibility for ethical action. We invite researchers to draw upon first person perspectives in design and STS approaches to acknowledge the limitations of individual control over technology development and to recognize opportunities to make change collectively. We believe these resources will help advance an actionable, relational approach to ethics that acknowledges the connection between emotion

and ethical practice while refraining from trivializing complex issues as matters of individual responsibility.

8 ACKNOWLEDGMENTS

REFERENCES

[1] AI-HLEG. 2019. *Ethics Guidelines for Trustworthy AI*. Text. European Commission. https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

[2] Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2022. Machine bias. In *Ethics of Data and Analytics*. Auerbach Publications, New York, US, 254–264.

[3] Maria Arnelid, Katherine Harrison, and Ericka Johnson. 2022. What Does It Mean to Measure a Smile? Assigning numerical values to emotions. *Valuation Studies* 9, 1 (Dec. 2022), 79–107. https://doi.org/10.3384/VS.2001-5992.2022.9.1.79-107

[4] Madeline Balaam, Rob Comber, Rachel E. Clarke, Charles Windlin, Anna Ståhl, Kristina Höök, and Geraldine Fitzpatrick. 2019. Emotion Work in Experience-Centered Design. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300832

[5] Stephanie Ballard, Karen M. Chappell, and Kristen Kennedy. 2019. Judgment Call the Game: Using Value Sensitive Design and Design Fiction to Surface Ethical Concerns Related to Technology. In *Proceedings of the 2019 on Designing Interactive Systems Conference* (San Diego, CA, USA) *(DIS '19)*. Association for Computing Machinery, New York, NY, USA, 421–433. https://doi.org/10.1145/3322276.3323697

[6] Jeffrey Bardzell and Shaowen Bardzell. 2013. What is "Critical" about Critical Design?. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. Association for Computing Machinery, New York, NY, USA, 3297–3306. https://doi.org/10.1145/2470654.2466451

[7] Shaowen Bardzell. 2010. Feminist HCI: Taking Stock and Outlining an Agenda for Design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '10)*. Association for Computing Machinery, New York, NY, USA, 1301–1310. https://doi.org/10.1145/1753326.1753521

[8] Saba Bazargan-Forward and Deborah Tollefsen (Eds.). 2020. *The Routledge Handbook of Collective Responsibility*. Routledge, New York, NY, USA. https://doi.org/10.4324/9781315107608

[9] Robert D. Benford and David A. Snow. 2000. Framing Processes and Social Movements: An Overview and Assessment. *Annual Review of Sociology* 26 (2000), 611–639. http://www.jstor.org/stable/223459

[10] Ruha Benjamin. 2019. *Race after technology: abolitionist tools for the new Jim code*. Polity, Medford, MA.

[11] Janne Mascha Beuthel. 2022. The Armor of a Researcher: A Critical, Material Engagement with Somatic Experiences. *Revista Diseña* 20, 3 (2022). https://doi.org/10.7764/disena.20.article.3

[12] Geoffrey C Bowker and Susan Leigh Star. 2000. *Sorting things out: Classification and its consequences*. MIT press, Cambridge, Massachusetts, USA.

[13] Karen L. Boyd and Katie Shilton. 2021. Adapting Ethical Sensitivity as a Construct to Study Technology Design Teams. *Proceedings of the ACM on Human-Computer Interaction* 5, GROUP (July 2021), 217. https://doi.org/10.1145/3463929

[14] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (Jan. 2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[15] Barry Brown, Susanne Bødker, and Kristina Höök. 2017. Does HCI Scale? Scale Hacking and the Relevance of HCI. *Interactions* 24, 5 (aug 2017), 28–33. https://doi.org/10.1145/3125387

[16] Joy Buolamwini and Timnit Gebru. 2018. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency (Proceedings of Machine Learning Research, Vol. 81)*, Sorelle A. Friedler and Christo Wilson (Eds.). PMLR, Nice, France, 77–91. https://proceedings.mlr.press/v81/buolamwini18a.html

[17] Kathryn Bushby, Jody Chan, Shawna Druif, Kim Ho, and Elizabeth Anne Kinsella. 2015. Ethical tensions in occupational therapy practice: A scoping review. *British Journal of Occupational Therapy* 78, 4 (2015), 212–221. https://doi.org/10.1177/0308022614564770 arXiv:https://doi.org/10.1177/0308022614564770

[18] Shruthi Sai Chivukula, Colin M. Gray, and Jason A. Brier. 2019. Analyzing Value Discovery in Design Decisions Through Ethicography. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300307

[19] Shruthi Sai Chivukula, Chris Rhys Watkins, Rhea Manocha, Jingle Chen, and Colin M. Gray. 2020. Dimensions of UX Practice that Shape Ethical Awareness. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376459

[20] Victoria Clarke and Virginia Braun. 2021. *Thematic analysis: a practical guide.* SAGE Publications Ltd. 1–100 pages.

[21] Mark Coeckelbergh. 2020. Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability. *Science and Engineering Ethics* 26, 4 (Oct. 2020), 2051–2068. https://doi.org/10.1007/s11948-019-00146-8

[22] Sasha Costanza-Chock. 2020. *Design justice: community-led practices to build the worlds we need.* The MIT Press, Cambridge, Massachusetts, USA.

[23] Kate Crawford and Vladan Joler. 2018. Anatomy of an AI System: The Amazon Echo As An Anatomical Map of Human Labor, Data and Planetary Resources. https://anatomyof.ai

[24] Maria Puig de La Bellacasa. 2017. *Matters of care: Speculative ethics in more than human worlds.* Vol. 41. U of Minnesota Press.

[25] Laura Devendorf, Kristina Andersen, and Aisling Kelliher. 2020. Making Design Memoirs: Understanding and Honoring Difficult Experiences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376345

[26] Catherine D'Ignazio and Lauren F. Klein. 2020. *Data feminism.* The MIT Press, Cambridge, Massachusetts, USA.

[27] Virginia Dignum. 2019. *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way.* Springer, Cham. https://doi.org/10.1007/978-3-030-30371-6

[28] Joseph Dumit. 2014. Writing the implosion: teaching the world one thing at a time. *Cultural Anthropology* 29, 2 (2014), 344–362.

[29] Arturo Escobar. 2018. *Designs for the Pluriverse: Radical Interdependence, Autonomy, and the Making of Worlds.* Duke University Press, New York, USA.

[30] Casey Fiesler. 2021. Innovating like an Optimist, Preparing like a Pessimist: Ethical Speculation and the Legal Imagination. *Colorado Technology Law Journal* 19, 1 (2021), 1–18.

[31] Clàudia Figueras, Harko Verhagen, and Teresa Cerratto Pargman. 2022. Exploring tensions in Responsible AI in practice: An interview study on AI practices in and for Swedish public organizations. *Scandinavian Journal of Information Systems* 34, 2 (2022), 6.

[32] Mary Flanagan, Daniel C. Howe, and Helen Nissenbaum. 2005. Values at Play: Design Tradeoffs in Socially-Oriented Game Design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Portland, Oregon, USA) *(CHI '05)*. Association for Computing Machinery, New York, NY, USA, 751–760. https://doi.org/10.1145/1054972.1055076

[33] Christopher Frauenberger, Marjo Rauhala, and Geraldine Fitzpatrick. 2017. In-Action Ethics. *Interacting with Computers* 29, 2 (2017), 220–236. https://doi.org/10.1093/iwc/iww024

[34] Batya Friedman. 1996. Value-Sensitive Design. *Interactions* 3, 6 (dec 1996), 16–23. https://doi.org/10.1145/242485.242493

[35] Batya Friedman and David Hendry. 2019. *Value Sensitive Design: Shaping Technology with Moral Imagination.* The MIT Press, Cambridge, Massachusetts, USA.

[36] Batya Friedman, Daniel C Howe, and Edward Felten. 2002. Informed consent in the Mozilla browser: implementing value-sensitive design. In *Proceedings of the 35th Annual Hawaii International Conference on System Sciences*. IEEE Comput. Soc, Big Island, HI, USA, 10. https://doi.org/10.1109/HICSS.2002.994366

[37] Batya Friedman, Peter H. Kahn Jr., Jennifer Hagman, Rachel L. Severson, and Brian Gill. 2006. The Watcher and the Watched: Social Judgments About Privacy in a Public Place. *Human–Computer Interaction* 21, 2 (2006), 235–272. https://doi.org/10.1207/s15327051hci2102_3

[38] Batya Friedman, Peter H. Khan, and Daniel C. Howe. 2000. Trust Online. *Commun. ACM* 43, 12 (dec 2000), 34–40. https://doi.org/10.1145/355112.355120

[39] Rachael Garrett, Kristina Popova, Claudia Núñez-Pacheco, Thorhildur Asgeirsdottir, Airi Lampinen, and Kristina Höök. 2023. Felt Ethics: Cultivating Ethical Sensibility in Design Practice. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 1, 15 pages. https://doi.org/10.1145/3544548.3580875

[40] Carol Gilligan. 1977. In a different voice: Women's conceptions of self and of morality. *Harvard educational review* 47, 4 (1977), 481–517.

[41] Lisa Gitelman (Ed.). 2013. *"Raw data" is an oxymoron.* The MIT Press, Cambridge, Massachusetts, USA.

[42] Colin M Gray and Shruthi Sai Chivukula. 2019. Ethical Mediation in UX Practice. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association.ie for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3290605.3300408

[43] Colin M. Gray, Yubo Kou, Bryan Battles, Joseph Hoggatt, and Austin L. Toombs. 2018. The Dark (Patterns) Side of UX Design. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3173574.3174108

[44] Colin M Gray, Austin L Toombs, Ann Light, and John Vines. 2018. Editorial: Ethics, Values, and Designer Responsibility. In *Design as a catalyst for change*, C Storni, K Leahy, M McMahon, P Lloyd, and E Bohemia (Eds.). DRS International Conference 2018, Limerick, Ireland. https://doi.org/10.21606/drs.2018.003

[45] Daniel Greene, Anna Lauren Hoffmann, and Luke Stark. 2019. Better, Nicer, Clearer, Fairer: A Critical Assessment of the Movement for Ethical Artificial Intelligence and Machine Learning. In *Proceedings of the 52nd Hawaii International Conference on System Sciences*. IEEE Computer Society Press, Hawaii, USA, 2122–2131. https://doi.org/10.24251/HICSS.2019.258

[46] Thilo Hagendorff. 2020. The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds and Machines* 30, 1 (2020), 99–120. https://doi.org/10.1007/s11023-020-09517-8

[47] Sandra Harding. 2015. *Objectivity and diversity. Another Logic of Scientific Research.* University of Chicago Press, Chicago, USA.

[48] Karey Helms. 2022. A Speculative Ethics for Designing with Bodily Fluids. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, Article 13, 11 pages. https://doi.org/10.1145/3491101.3516395

[49] Arlie Russel Hochschild. 2012. *The managed heart: commercialization of human feeling* (updated ed ed.). University of California Press, Berkeley, Calif. London.

[50] James Hodge, Sarah Foley, Rens Brankaert, Gail Kenning, Amanda Lazar, Jennifer Boger, and Kellie Morrissey. 2020. Relational, Flexible, Everyday: Learning from Ethics in Dementia Research. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–16. https://doi.org/10.1145/3313831.3376627

[51] David Holman, David Martinez-Iñigo, and Peter Totterdell. 2008. Emotional Labour and Employee Well-being: An Integrative Review. In *Research Companion to Emotion in Organizations*, Neal M Ashkanasy and Cary L Cooper (Eds.). Edward Elgar Publishing, Cheltenham, UK, Chapter 18, 301–316.

[52] Kristina Höök. 2018. *Designing with the body: Somaesthetic interaction design.* MIT Press, Cambridge, MA, USA.

[53] Kristina Höök, Baptiste Caramiaux, Cumhur Erkut, Jodi Forlizzi, Nassrin Hajinejad, Michael Haller, Caroline Hummels, Katherine Isbister, Martin Jonsson, George Khut, Lian Loke, Danielle Lottridge, Patrizia Marti, Edward Melcer, Florian Müller, Marianne Petersen, Thecla Schiphorst, Elena Segura, Anna Ståhl, Dag Svanæs, Jakob Tholander, and Helena Tobiasson. 2018. Embracing First-Person Perspectives in Soma-Based Design. *Informatics* 5, 1 (Feb. 2018), 8. https://doi.org/10.3390/informatics5010008

[54] Kristina Höök, Sara Eriksson, Marie Louise Juul Søndergaard, Marianela Ciolfi Felice, Nadia Campo Woytuk, Ozgun Kilic Afsar, Vasiliki Tsaknaki, and Anna Ståhl. 2019. Soma Design and Politics of the Body. In *Proceedings of the Halfway to the Future Symposium 2019* (Nottingham, United Kingdom) *(HTTF 2019)*. Association for Computing Machinery, New York, NY, USA, Article 1, 8 pages. https://doi.org/10.1145/3363384.3363385

[55] Noura Howell, Audrey Desjardins, and Sarah Fox. 2021. Cracks in the Success Narrative: Rethinking Failure in Design Research through a Retrospective Trioethnography. *ACM Trans. Comput.-Hum. Interact.* 28, 6, Article 42 (nov 2021), 31 pages. https://doi.org/10.1145/3462447

[56] Anna Jobin, Marcello Ienca, and Effy Vayena. 2019. The global landscape of AI ethics guidelines. *Nature Machine Intelligence* 1, 9 (Sept. 2019), 389–399. https://doi.org/10.1038/s42256-019-0088-2

[57] Laura Kalbag and Aral Balkan. 2017. Ethical Design Manifesto by ind.ie. https://ind.ie/ethical-design/

[58] Pratyusha Kalluri. 2020. Don't ask if artificial intelligence is good or fair, ask how it shifts power. *Nature* 583, 7815 (July 2020), 169–169. https://doi.org/10.1038/d41586-020-02003-2

[59] Emre Kazim and Adriano Soares Koshiyama. 2021. A high-level overview of AI ethics. *Patterns* 2, 9 (Sept. 2021), 100314. https://doi.org/10.1016/j.patter.2021.100314

[60] Carly Kind. 2020. The term 'ethical AI' is finally starting to mean something. https://venturebeat.com/2020/08/23/the-term-ethical-ai-is-finally-starting-to-mean-something/

[61] Max Liboiron. 2021. *Pollution Is Colonialism.* Duke University Press. https://doi.org/10.2307/j.ctv1jhvnk1

[62] Max Liboiron, Emily Simmonds, Edward Allen, Emily Wells, Jessica Melvin, Alex Zahara, and Charles Mather. 2021. Doing Ethics with Cod. In *Making & Doing: Activating STS through Knowledge Expression and Travel*. The MIT Press. https://doi.org/10.7551/mitpress/11310.003.0009

[63] Ann Light, Alison Powell, and Irina Shklovski. 2017. Design for Existential Crisis in the Anthropocene Age. In *Proceedings of the 8th International Conference on Communities and Technologies (C&T '17)*. Association for

Computing Machinery, New York, NY, USA, 270–279. https://doi.org/10.1145/3083671.3083688

[64] Sharon Lindberg, Petter Karlström, and Sirkku Männikkö Barbutiu. 2021. Design Ethics in Practice - Points of Departure. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 130 (apr 2021), 19 pages. https://doi.org/10.1145/3449204

[65] Michael A. Madaio, Luke Stark, Jennifer Wortman Vaughan, and Hanna Wallach. 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* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376445

[66] Jacob Metcalf, Emanuel Moss, and danah boyd. 2019. Owning Ethics: Corporate Logics, Silicon Valley, and the Institutionalization of Ethics. *Social Research: An International Quarterly* 86, 2 (2019), 449–476. https://muse.jhu.edu/article/732185

[67] Brent Mittelstadt. 2019. Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence* 1, 11 (Nov. 2019), 501–507. https://doi.org/10.1038/s42256-019-0114-4

[68] Shakir Mohamed, Marie-Therese Png, and William Isaac. 2020. Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence. *Philosophy & Technology* 33, 4 (2020), 659–684.

[69] Mike Monteiro and Vivianne Castillo. 2019. *Ruined by design: how designers destroyed the world, and what we can do to fix it.* Mule Design, Fresno.

[70] Jessica Morley, Luciano Floridi, Libby Kinsey, and Anat Elhalal. 2020. From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices. *Science and Engineering Ethics* 26, 4 (Aug. 2020), 2141–2168. https://doi.org/10.1007/s11948-019-00165-5

[71] Chris Nodder. 2013. *Evil by design: Interaction design to lead us into temptation.* John Wiley & Sons, Indianapolis, USA.

[72] Nel Noddings. 2013. *Caring: A Relational Approach to Ethics and Moral Education* (2 ed.). University of California Press, Berkeley, California. http://www.jstor.org/stable/10.1525/j.ctt7zw1nb

[73] Giovanna Nunes Vilaza, Kevin Doherty, Darragh McCashin, David Coyle, Jakob Bardram, and Marguerite Barry. 2022. A Scoping Review of Ethics Across SIGCHI. In *Designing Interactive Systems Conference (DIS '22)*. Association for Computing Machinery, New York, NY, USA, 137–154. https://doi.org/10.1145/3532106.3533511

[74] Ziad Obermeyer, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. 2019. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 366, 6464 (Oct. 2019), 447–453. https://doi.org/10.1126/science.aax2342

[75] The IEEE Global Initiative on Ethics of Autonomous and Intelligent System. 2019. *Ethically Aligned Design: A vision for prioritizing human well-being with autonomous and intelligent systems, First edition.* Technical Report. IEEE. https://ethicsinaction.ieee.org/

[76] Victor J. Papanek. 1985. *Design for the real world: human ecology and social change* (2nd ed., completely rev ed.). Academy Chicago, Chicago, Ill.

[77] Samir Passi and Steven J. Jackson. 2018. Trust in Data Science: Collaboration, Translation, and Accountability in Corporate Data Science Projects. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (Nov. 2018), 1–28. https://doi.org/10.1145/3274405

[78] Samir Passi and Phoebe Sengers. 2020. Making data science systems work. *Big Data & Society* 7, 2 (2020), 2053951720939605. https://doi.org/10.1177/2053951720939605 arXiv:https://doi.org/10.1177/2053951720939605

[79] Ajit G. Pillai, Thida Sachathep, and Naseem Ahmadpour. 2022. Exploring the Experience of Ethical Tensions and the Role of Community in UX Practice. In *Nordic Human-Computer Interaction Conference* (Aarhus, Denmark) *(NordiCHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 60, 13 pages. https://doi.org/10.1145/3546155.3546683

[80] Clémence Pinel, Barbara Prainsack, and Christopher McKevitt. 2020. Caring for data: Value creation in a data-intensive research laboratory. *Social Studies of Science* 50, 2 (2020), 175–197. https://doi.org/10.1177/0306312720906567 PMID: 32053062.

[81] Marie-Therese Png. 2022. At the Tensions of South and North: Critical Roles of Global South Stakeholders in AI Governance. In *2022 ACM Conference on Fairness, Accountability, and Transparency*. Association for Computing Machinery, New York, NY, USA, 1434–1445. https://doi.org/10.1145/3531146.3533200

[82] Kristina Popova, Rachael Garrett, Claudia Núñez Pacheco, Airi Lampinen, and Kristina Höök. 2022. Vulnerability as an Ethical Stance in Soma Design Processes. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 178, 13 pages. https://doi.org/10.1145/3491102.3501994

[83] Alison B Powell, Funda Ustek-Spilda, Sebastián Lehuedé, and Irina Shklovski. 2022. Addressing ethical gaps in 'Technology for Good': Foregrounding care and capabilities. *Big Data & Society* 9, 2 (2022), 20539517221113774. https://doi.org/10.1177/20539517221113774 arXiv:https://doi.org/10.1177/20539517221113774

[84] Bogdana Rakova, Jingying Yang, Henriette Cramer, and Rumman Chowdhury. 2021. Where Responsible AI Meets Reality: Practitioner Perspectives on Enablers for Shifting Organizational Practices. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 7 (apr 2021), 23 pages. https://doi.org/10.1145/3449081

[85] Samar Sabie and Tapan Parikh. 2019. Cultivating Care through Ambiguity: Lessons from a Service Learning Course. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300507

[86] Nithya Sambasivan, Erin Arnesen, Ben Hutchinson, Tulsee Doshi, and Vinodkumar Prabhakaran. 2021. Re-imagining Algorithmic Fairness in India and Beyond. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*. Association for Computing Machinery, New York, NY, USA, 315–328. https://doi.org/10.1145/3442188.3445896

[87] Filippo Santoni de Sio and Giulio Mecacci. 2021. Four Responsibility Gaps with Artificial Intelligence: Why they Matter and How to Address them. *Philosophy & Technology* 34, 4 (Dec. 2021), 1057–1084. https://doi.org/10.1007/s13347-021-00450-x

[88] Daniel Schiff, Bogdana Rakova, Aladdin Ayesh, Anat Fanti, and Michael Lennon. 2021. Explaining the Principles to Practices Gap in AI. *IEEE Technology and Society Magazine* 40, 2 (June 2021), 81–94. https://doi.org/10.1109/MTS.2021.3056286 Conference Name: IEEE Technology and Society Magazine.

[89] Nick Seaver. 2021. Care and Scale: Decorrelative Ethics in Algorithmic Recommendation. *Cultural Anthropology* 36, 3 (Aug. 2021), 509–537. https://doi.org/10.14506/ca36.3.11 Number: 3.

[90] Phoebe Sengers, Kirsten Boehner, Shay David, and Joseph 'Jofish' Kaye. 2005. Reflective Design. In *Proceedings of the 4th Decennial Conference on Critical Computing: Between Sense and Sensibility* (Aarhus, Denmark) *(CC '05)*. Association for Computing Machinery, New York, NY, USA, 49–58. https://doi.org/10.1145/1094562.1094569

[91] Katie Shilton. 2013. Values Levers: Building Ethics into Design. *Science, Technology, & Human Values* 38, 3 (May 2013), 374–397. https://doi.org/10.1177/0162243912436985

[92] Irina Shklovski and Carolina Némethy. 2022. Nodes of certainty and spaces for doubt in AI ethics for engineers. *Information, Communication & Society* 0, 0 (Jan. 2022), 1–17. https://doi.org/10.1080/1369118X.2021.2014547

[93] Robert Soden, David Ribes, Seyram Avle, and Will Sutherland. 2021. Time for Historicism in CSCW: An Invitation. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 459 (oct 2021), 18 pages. https://doi.org/10.1145/3479603

[94] Katta Spiel. 2018. *Evaluating Experiences of Autistic Children with Technologies in Co-Design.* Doctor of Technical Sciences. Vienna University of Technology (TU Wien), Vienna, Austria. Publisher: Gesellschaft für Informatik eV.

[95] Katta Spiel, Emeline Brulé, Christopher Frauenberger, Gilles Bailly, and Geraldine Fitzpatrick. 2018. Micro-Ethics for Participatory Design with Marginalised Children. In *Proceedings of the 15th Participatory Design Conference: Full Papers - Volume 1* (Hasselt and Genk, Belgium) *(PDC '18)*. Association for Computing Machinery, New York, NY, USA, Article 17, 12 pages. https://doi.org/10.1145/3210586.3210603

[96] Anna Ståhl, Vasiliki Tsaknaki, and Madeline Balaam. 2021. Validity and Rigour in Soma Design-Sketching with the Soma. *ACM Transactions on Computer-Human Interaction (TOCHI)* 28, 6 (2021), 1–36.

[97] Luke Stark. 2019. Facial Recognition is the Plutonium of AI. *XRDS* 25, 3 (apr 2019), 50–55. https://doi.org/10.1145/3313129

[98] Angelika Strohmayer, Mary Laing, and Rob Comber. 2017. Technologies and Social Justice Outcomes in Sex Work Charities: Fighting Stigma, Saving Lives. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 3352–3364. https://doi.org/10.1145/3025453.3025615

[99] Norman Makoto Su, Amanda Lazar, and Lilly Irani. 2021. Critical Affects: Tech Work Emotions Amidst the Techlash. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 179 (apr 2021), 27 pages.

[100] Joan C. Tronto. 1993. *Moral boundaries: a political argument for an ethic of care.* Routledge, New York, USA.

[101] Anna Lowenhaupt Tsing. 2015. *The Mushroom at the End of the World: On the Possibility of Life in Capitalist Ruins.* Princeton University Press.

[102] Ibo van de Poel, Lambèr Royakkers, and Sjoerd D. Zwart. 2015. *Moral Responsibility and the Problem of Many Hands* (0 ed.). Routledge, New York, USA. https://doi.org/10.4324/9781315734217

[103] Francisco J. Varela. 1999. *Ethical know-how: action, wisdom, and cognition.* Stanford University Press, Stanford, Calif.

[104] Michael Veale, Max Van Kleek, and Reuben Binns. 2018. Fairness and Accountability Design Needs for Algorithmic Support in High-Stakes Public Sector Decision-Making. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3173574.3174014

[105] Jessica Vitak, Katie Shilton, and Zahra Ashktorab. 2016. Beyond the Belmont principles: Ethical challenges, practices, and beliefs in the online data research community. In *Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing*. ACM, San Francisco, USA, 941–953.

[106] Sandra Wachter, Brent Mittelstadt, and Luciano Floridi. 2017. Transparent, explainable, and accountable AI for robotics. *Science Robotics* 2, 6 (may 2017), eaan6080. https://doi.org/10.1126/scirobotics.aan6080

[107] David Gray Widder, Derrick Zhen, Laura Dabbish, and James Herbsleb. 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* (Chicago, IL, USA) *(FAccT '23)*. Association for Computing Machinery, New York, NY, USA, 467–479. https://doi.org/10.1145/3593013.3594012

[108] Langdon Winner. 1980. Do Artifacts Have Politics? *Daedalus* 109, 1 (1980), 121–136. http://www.jstor.org/stable/20024652

[109] Richmond Y. Wong. 2021. Tactics of Soft Resistance in User Experience Professionals' Values Work. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 355 (oct 2021), 28 pages. https://doi.org/10.1145/3479499

[110] Richmond Y. Wong, Michael A. Madaio, and Nick Merrill. 2023. Seeing Like a Toolkit: How Toolkits Envision the Work of AI Ethics. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW1, Article 145 (apr 2023), 27 pages. https://doi.org/10.1145/3579621

[111] Haiyi Zhu, Bowen Yu, Aaron Halfaker, and Loren Terveen. 2018. Value-Sensitive Algorithm Design: Method, Case Study, and Lessons. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (Nov. 2018), 1–23. https://doi.org/10.1145/3274463

[112] Jarrett Zigon. 2007. Moral breakdown and the ethical demand: A theoretical framework for an anthropology of moralities. *Anthropological Theory* 7, 2 (2007), 131–150. https://doi.org/10.1177/1463499607077295