



# "Why do we do this?": Moral Stress and the Affective Experience of Ethics in Practice

Sonja Rattay  
University of Copenhagen  
Copenhagen, Denmark  
ITU Austria  
Linz, Austria  
srr@di.ku.dk

Ville Vakkuri  
University of Vaasa  
Vaasa, Finland  
ville.vakkuri@uwasa.fi

Marco C. Rozendaal  
Industrial Design Engineering  
Delft University of Technology  
Delft, South-Holland, Netherlands  
m.c.rozendaal@tudelft.nl

Irina Shklovski  
University of Copenhagen  
Copenhagen, Denmark  
Linköping University  
Linköping, Sweden  
ias@di.ku.dk

## Abstract

A plethora of toolkits, checklists, and workshops have been developed to bridge the well-documented gap between AI ethics principles and practice. Yet little is known about effects of such interventions on practitioners. We conducted an ethnographic investigation in a major European city organization that developed and works to integrate an ethics toolkit into city operations. We find that the integration of ethics tools by technical teams destabilises their boundaries, roles, and mandates around responsibilities and decisions. This lead to emotional discomfort and feelings of vulnerability, which neither toolkit designers nor the organization had accounted for. We leverage the concept of moral stress to argue that this affective experience is a core challenge to the successful integration of ethics tools in technical practice. Even in this best case scenario, organisational structures were not able to deal with moral stress that resulted from attempts to implement responsible technology development practices.

## CCS Concepts

• **Social and professional topics** → **Codes of ethics**; • **Human-centered computing** → **Empirical studies in HCI**.

## Keywords

Moral stress, ethics in practice, responsible technology development, affective experience

### ACM Reference Format:

Sonja Rattay, Ville Vakkuri, Marco C. Rozendaal, and Irina Shklovski. 2025. "Why do we do this?": Moral Stress and the Affective Experience of Ethics in Practice. In *CHI Conference on Human Factors in Computing Systems (CHI '25)*, April 26–May 01, 2025, Yokohama, Japan. ACM, New York, NY, USA, 15 pages. <https://doi.org/10.1145/3706598.3713264>



This work is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License.

*CHI '25, Yokohama, Japan*

© 2025 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-1394-1/25/04

<https://doi.org/10.1145/3706598.3713264>

## 1 Introduction

Debates about ethics in computing are decidedly not new, but the proliferation of complex data-driven systems has resulted in an explosion of scholarly work and regulatory efforts. Research has circled around defining high-level ethical principles [29, 31], shifting between a focus on ethics and a discussion of values [33] and developing a range of guidelines [20, 42, 48] and codes [7]. These efforts fostered development of principles-based checklists [98], toolkits [64], workshops [6, 82, 86], and a variety of materials designed to support reflection in technology design and development [21, 39, 40]. The multitude of publications and materials mainly highlight that so far, there is little consensus on how to reliably address ethical challenges posed by rapid technological progress.

Where such efforts have been studied, the prominent finding has been a lack of impact [59, 96] and warnings that principles, while important, are difficult to effectively translate into practice [62]. Researchers have investigated the difficulties faced by practitioners, highlighting a mismatch between expectations of operationalised ethics and the messy reality of everyday work life [16, 38, 52, 94, 95]. Yet despite the emerging emphasis on understanding how ethics is done in practice [52, 57, 71, 97], little research has been able to investigate how integration of ethics toolkits and other interventions may impact technological development practices long term.

In this study we ask how do practitioners experience the implementation of an ethics toolkit in an everyday organisational context? We conducted an ethnographic study with a technical team in a major European city. This complex public services organization has already developed its own internal principles-based ethics toolkit, integrating it into city operations that focus on the design, development, and implementation of data driven technologies (we will use the term "design of AI systems" to refer to this process throughout the paper). Rather than evaluating the toolkit itself, we focus on what such an implementation does to the practitioners that take it on board. We find that practitioners in our study must find ways to cope with uncomfortable feelings, such as anxiety and frustration, that arise from identifying and engaging with challenging moral situations that become more visible as they use the

toolkit. We leverage the concept of moral stress [73] from moral psychology to analyze this affective experience of ethics in practice and to understand the emotional labor that affects the outcomes of ethics toolkit implementation in organisations. We identify three relational configurations relevant for the affective experience of ethics in practice – within team, beyond team, and beyond organization and showcase how moral stress unfolds differently across these.

The paper makes the following contributions:

1. We demonstrate how the integration of ethics tools destabilises established boundaries, roles, and mandates around responsibilities and decisions in an organizational context, introducing greater uncertainty about what constitutes the right action.
2. We introduce the concept of moral stress as an unavoidable yet completely overlooked affective cost of ethics interventions for practitioners, which can explain why so many efforts towards responsible technology development fail to reach intended outcomes.
3. We illustrate the role of organizational structures and the importance of well-functioning teams as key determinants for the design of effective interventions towards responsible technology development

## 2 Related Work

### 2.1 The Theory-Practice Gap in AI Ethics

The gap between AI ethics principles and practice is well documented [29, 62] and there is consensus that crossing this gap is a non-trivial problem [65]. For example, in a review of 200 AI ethics guidelines and recommendations, Corrêa et al. [20] find that prescriptive normative claims are typically presented without considerations for how to achieve them. While there seems to be a convergence around which principles are most important, the interpretation and justifications of why and how these principles matter diverge widely [48]. Even the ACM code of ethics, while a decidedly important document, has little apparent impact on professionals in the tech community [59]. In a scoping review of responsible AI guidelines, Sadek et al. [75] note that the abstract nature of the principles and the lack of clear implementation procedures are important reasons for why the gap between principles and practice persists.

In parallel to prescriptive principle-based approaches, we see a variety of materials targeting ethical concerns in technical practice, such as games [2, 6], card-sets [4, 13, 88, 90], toolkits [36, 85, 101], and workshops [1, 26, 44, 67] that have been produced both by researchers and industry actors to operationalize ethical considerations into development processes. In the following we refer to such materials as ethics interventions. Reviews of these interventions consistently show that they tend to fall short of impact in the AI industry [15, 42, 99] as they do not account for the environments in which such tech practices are situated [16, 51, 96, 100]. For example, Morley and colleagues [64] demonstrate that almost all existing translational tools and methods are either too flexible (vulnerable to ethics washing) or too strict (unresponsive to context). The decoupling of policies, practices, and outcomes leads to practitioners facing major hurdles when trying to integrate AI ethics practices into development processes and organizational structures [2]. Interviews with industry leaders in particular report

that lacking awareness and feasible practices are a challenge [53], and that ethics efforts are perceived to be in tension with industry structures and values such as technological solutionism and market fundamentalism [53, 60].

### 2.2 Ethics as a social and relational practice

Where principles-based approaches often take their departure in philosophical considerations of ethical concerns, studies of ethics in practice in technical contexts demonstrate that technical practitioners are already ethically engaged through other means [23, 25, 45, 86, 95, 97]. Shilton's [77] notion of value levers for example describes how organizational and operational structures in design teams and labs influence the way values and ethical deliberations are being thematized, normalized, and included or excluded in daily routines. Shilton described different infrastructural aspects that support and embed discussions about values in the design of technology, calling attention to the fact that in a market driven design field, the constant pressure of technical innovation makes it more difficult for teams to make the time for a slow and deliberate value-driven design process. More recently, Lindberg, Karlström and Barbutiu [52] report on the persistent organisational barriers that make it difficult to create emotional buy in from practitioners to invest time and resources on ethics efforts, in particular where such efforts take more complex forms to encourage practitioners to reflect and engage with ethical implications of their work beyond "tick box exercises".

Such research highlights that ethics in practice is not the same as ethics as normative inquiry. The modularized nature of the software development workflow, inherent to any AI development, makes it difficult to pin down where exactly accountabilities and responsibilities for principles and values are located among the developers and other stakeholders [94]. Organisational processes such as project planning, resource management, and team hierarchies strongly influence ethical decision-making dynamics [24]. The central debate in tech ethics is therefore not whether or which ethics is desirable, but what "ethics" entails, who gets to define it, and what kind of impact those questions have in practice [41]. Wong and colleagues [99] note that ethics toolkits often lack guidance around how to navigate labor, organizational, and institutional power dynamics as they relate to technical work. Currently, the gap between principles and practice requires a lot of invisible work from technologists to navigate hierarchies, market dynamics, and the lack of awareness in the rest of the organization, which mostly goes unacknowledged [23, 92].

Demands and idealistic notions of ethical reflection and action then clash with the reality of everyday industry practice, in which tech logics and organisational structures stand in stark opposition to the necessary resources required for ethical reflection - such as uncertainty, criticality, and slowness [81]. Yet little research has explored the impact and affective experience of integrating ethics tools into environments where such clashes are inevitable. We follow scholars of ethics in practice [27, 71, 72, 79] who argue against a too individualized approach to ethics, and consider ethics as a relational practice, where responsibility lies between the technical expert and the organizational structures within which they operate [50]. We also build on research that argues for considering ethics

as a situated and lived experience [78], suggesting that the emotional context of ethics is important for bridging the gap between principles and practice [28]. Prior research that reports on concrete methods, tools, and specific case studies of their usage, such as value dams [61], and a variety of similar ethics interventions developed in HCI, tell us how practitioners deal with concrete trade offs and argue that those skills are transferable to future projects. They do not report on the longer term shifts in mindset and how that makes affected practitioners feel about their work moving forward. We therefore investigate ethical considerations as emotional and embodied [83], especially where they challenge existing social and political norms in an organisational context, coming into conflict with the techno-optimism that often accompanies technology development [37].

### 2.3 Emotional response, ethical sensitivity and moral stress

Moral reasoning and ethical decision-making have an emotional component that has long been overlooked by responsible AI research efforts. Deep reflection that is required in design is a type of emotion work, that can come at the cost of feelings of guilt, self-blame, and emotional exhaustion [5]. Exploring ethical considerations in design through a soma design perspective, Garrett et al. [35] and Popova et al. [70] demonstrate that connecting ethical sensitivity to emotional experience can provide a generative approach to understanding the embodied perspective of ethical decision making, but also involves distancing and vulnerability as key emotional dimensions [69]. Investigating an applied setting, Ma and colleagues report on the self-doubts, insecurities, and suffering faced by designers being limited in their decision-making power in addressing unfair business models through design decisions [56]. Engaging with ethical considerations then, can be a challenging and vulnerable, emotion-laden process, but so far HCI has lacked the language to systematically describe the experience and its implications.

Outside of HCI, the challenges of moral conduct and ethics in practice have been well studied from a psychological perspective. In fields such as management studies and nursing studies the fact that practitioners respond emotionally to moral quandaries and ethical dilemmas is established consensus. Studies from these fields provide us with defined vocabulary to identify internal processes around moral decision-making in practice. Jordan [49] for example defines the trait of having developed a situated attentiveness for the moral relevance of decisions and actions, in particular in connection to specific values, as *moral awareness*. Based on a review of 108 studies of from fields such as medicine and accounting, Boyd and Shilton [10] review the concept of *ethical sensitivity* for its relevance for technical practice. They conceptualise it as a three step process comprised of recognition (noticing an ethical problem, due to moral awareness), particularisation (understanding the situation that creates it) and judgment (being able to decide what to do about it).

Moral awareness and ethical sensitivity describe important parts of moral reasoning that precede morally motivated action and provide vocabulary to analyse the conduct and experiences of practitioners grappling with ethical challenges. An increase in ethical

sensitivity and moral awareness is a common goal for practical ethics interventions in tech [9]. The assumption is that expanding the worldview of developers and designers will lead to engagement in deeper ethical reflection, which can result in increased ethical sensitivity, in turn leading to morally sound development efforts and responsible technology design [91].

However, increased ethical sensitivity and moral awareness can also have negative consequences [93]. An increase of self-reflective behaviour can make people more self-conscious and uncertain about their own actions, and a higher moral awareness is linked to higher levels of stress when roles and environment do not align [3]. Research from nursing studies reports that nurses suffer when the constraints of nursing work, such as the lack of time, supervisory reluctance or institutional policy, limit or prevent capacity for moral action [19]. In management studies Reynolds and colleagues [74] identify suffering as a result of having to make morally relevant decisions, especially when there is a discrepancy between individual and organizational goals or views of the right actions. The discrepancy can come from having to choose between what are seen as right and wrong actions, but it can also occur when choosing between two legitimate but contradictory right options or two that appear equally wrong [74].

These adverse experiences as outcomes of increased ethical sensitivity are summarised by the concept of *moral stress* [73] (in nursing studies termed moral distress [46, 55, 63]). Reynolds and colleagues define moral stress as "a psychological state (both cognitive and emotional) marked by anxiety and unrest due to an individual's uncertainty about his or her ability to fulfill relevant moral obligations." [74]. This type of stress arises from having to manage a mismatch between moral ideals and reality [73, 74]. Emotionally, moral stress manifests outwardly through anxiety, frustration, and anger [55]. While there is disagreement about the precise conditions that cause moral stress, work load, time pressure, and role ambiguity are cited as the most common antecedents [46, 63]. When unaddressed, moral stress is reported to lower the quality of patient care or management decision-making, resulting in avoidance of morally relevant situations and even leading people to change workplace or occupation [63]. Research suggests a range of organisational mitigation strategies in order to avoid cognitive and emotional burn out [73].

Despite the richness of research looking into operationalising ethical considerations into technical practice tools, few studies have explored the *lived experience* of such operationalisation efforts. While the importance of emotions in ethical decision making is starting to be explored in HCI research, no conceptual bridge yet exists between operationalised ethics efforts and the affective experiences they trigger, especially where such experiences are negative. Ethics toolkits and similar materials to operationalize ethics are often expected to function like any other organizational process, while the surrounding power structures and required relational labor are overlooked [99]. In addition, where ethical sensitivity has been researched in the field of technology design, it has been as a property of individuals, while technology design typically happens in teams and collaborative structures [10].

We present insights from an ethnographic study of technology development processes that have integrated ethics tools into everyday practice, paying attention to the embodied, situated, and

emotional context of doing ethics with such tools. Leveraging the extant research on moral stress, we propose it as a valuable concept to provide vocabulary to help interpret our observations of ethical deliberation beyond acts of reflection, towards understanding the lived affective experience of practitioners. We see moral stress as an additional dimension contributing to our understanding of why well-intended ethics efforts struggle with sustainable adoption within industry.

### 3 Research Context

The effects of rapid digitalisation of the public sector launched a debate in Europe about the implications and ethical challenges of smart data-driven technologies in general and of their impact on the public domain in particular [30, 87]. Multiple major European cities have responded by developing codes of conduct, principles, value sets, and manifestos to inform and guide the future development of data driven systems. The London Office of Technology and Innovation, a coalition of the GLA, the London Councils and the London borough councils, for example, established the London Data Ethics Service [66], with in-house experts providing project facilitation, organisational development, and pan-London policy and research. Amsterdam developed the prominent Tada Manifesto [17], in collaboration with citizens, businesses, and governmental bodies [18]. Barcelona launched the Ethical Digital Standards initiative [11], which includes digital service standards, a Technology Code of Conduct, a set of methodologies for agile development in line with these standards, and the Ethical Digital Standard Manifesto [12], which outlines the cities vision of technological sovereignty and digital rights for cities.

Our study emerged out of a rare opportunity to observe the integration of such ethics tools into technical practice within a complex public services organization. We conducted an ethnographic investigation in collaboration with a European city that has launched their own responsible data manifesto and invested major efforts over the last five years to align their innovation agenda with this manifesto. The municipal organization brought together a team to develop their own version of a workshop format and an accompanying toolkit for it and to help integrate these tools into city operations. To ensure anonymity for our participants, we will refer to the city as the City, and will only describe the framework structure to provide context for our observations. In order to give the reader context, we will explain how we encountered and understood the ethics toolkit and its components. Rather than evaluating the ethics toolkit itself, we focus on the teams reactions to and experiences of the process of integrating the toolkit, as well as the social, emotional and organisational effects that the adoption of the responsible data manifesto had. We will introduce the team<sup>1</sup> and their roles, before we discuss their experiences with integrating the toolkit into their work.

#### 3.1 The Manifesto and The Toolkit

In response to a deep concern about the impact of technology on the world, the City chose the form of a manifesto to express and communicate their principles. More than a code of conduct or a

set of principles, a manifesto calls for collaborative and pro-active claiming of responsibility [34]. In addition to providing guidance, manifestos are meant to be persuasive, to demand attention, and to mobilize [54]. The City Manifesto names the concerns and uncertainties as “the big questions of the digital revolution”, calls for the forming of a community, and assigns responsibility to this community in order to make the City a moral thought leader and an example for how to address these big questions well. The Manifesto was developed in a broad collaboration with citizens, businesses, and governmental bodies, and provides a list of principles for the responsible use of data in the development of data driven technologies. It is accompanied by a wide variety of materials for distribution, such as posters, a website, and other promotional material. It signals the commitment of the City to addressing the challenges of developing smart city technologies responsibly and democratically. The City and involved parties consider the Manifesto a successful result of a complex, collaborative, and inclusive process. Over the last five years, the City has worked on implementing the Manifesto into their internal technology development processes, producing multiple initiatives to address and fulfil the expressed values. Recognising that the translation of generally formulated principles into concrete and practical action demands resources such as time, skill, and effort, the City committed to developing the required materials and competences internally.

The work done by the City to follow through on their commitment to the principles of the Manifesto is a positive example of a public organisation working to close the gap between theory and practice in tech ethics. Over the last five years, through multiple initiatives, the City brought together an in-house team of ethics experts and designers to create materials and support that enable teams to translate the principles of the Manifesto into their work practices. One of the outcomes is a Toolkit developed by the team of in-house ethics experts and designers, specifically tailored to the work structure of the City’s technical teams. The Toolkit consists of workshop materials designed to help teams develop a sensitivity for how ethical values are connected to their own everyday work and is currently slowly being adopted by project teams throughout the organisation.

#### 3.2 The Toolkit workshops

At the beginning of our fieldwork with the City, we met the group involved in the creation and management of the ethics Toolkit. Tanya was one of the ethics experts who was deeply involved in the development of the Manifesto prior to joining the City. Now, she is part of the City as an ethics expert with the mandate to help integrate the Manifesto into the City operations through the Toolkit. The work of integrating the principles into the City DNA, as Tanya frames it, has been a slow but steady process. Tanya’s team has created videos and materials for employees to access, but the most effective way to become familiar with the Manifesto is to book a training workshop with Tanya or one of the expert facilitators. In our first in-person meeting at the beginning of the fieldwork, Tanya explains the intentions behind the workshop:

*“We worked out methods so that teams, when they get the time to think about ethical implications, they can do this in a structured manner, that it’s replicable for*

<sup>1</sup>For the sake of a more personal read, we have given the team members names, however, all names used in this paper are fictional.

*several teams and projects, so they could kind of use the same method [in different projects]. And the results are shareable, they begin to help build this body of knowledge other people can learn from, that can be shared outside the organization [...]. And, especially, how to make sure that the theoretical idea of thinking about values becomes something that's workable, and something that seems [...] like they enjoy working [with it and feel like it] improves their work."*

The training workshop that Tanya's team offers is modular. When a team books a training workshop with an ethics expert for the first time, the first part is a general introduction to considering ethical implications when working with data. Tanya explains that she often has to start with the basics, as some people have no awareness of ethical issues at all within their work. This part is often tailored to the specific needs of the teams. After this introduction, the teams go through exercises that dive into the ethical implications of their specific project. Tanya says she considers it important to separate the introduction and the project-related part, because she wants to put the teams in charge when working through the project-related exercises of the workshop.

The materials for the project-related exercises include a huge poster, that spreads out like a tablecloth, putting the necessary conversations visibly on the table. The workshop format consists of four steps:

- (1) Find the "North Star" asks the teams to specify what their project is about, and for whom. Tanya reports that at this step, teams often encounter unexpected results, even when they have been working together for a long time already.
- (2) The team goes through questions that guide them through the process of connecting the Manifesto values to their project.
- (3) The teams rate themselves on how they are doing on each value, and list the risks associated with lower scores in a diagram.
- (4) After the teams have the risks of their project listed in the diagram, they brainstorm ideas on how to mitigate them, and then prioritize mitigation strategies. The teams formulate the outcomes of this process as a project brief for themselves. The brief is the final part of the big poster, and is designed to be cut off, so that the team can hang it up at their desks and keep handy for checking again later.

After the training workshop, the expectation is that the teams will run their own smaller version of the workshop for each new project and revisit the methods and questions throughout the projects. The materials for each of the four steps are always available to the teams online so that they can repeat the exercises as desired. Tanya recommends that at the end of the training workshop, the teams choose a one member to take on the role of the internal *Ethics Owner* [60].<sup>2</sup>, who is in charge of ensuring the ongoing engagement with the principles, either by setting up and facilitating recurring workshops themselves or by delegating it to other team members.

<sup>2</sup>The team uses a different term to describe this role, however for anonymisation purposes we use the term Ethics Owner, defined by [60], as it is similar in meaning.

## 4 Data collection and analysis

We were invited to spend time with technical teams at the innovation department within city operations in the fall of 2023, when the manifesto had been around for 5 years, and the developed workshop materials had been used for about a year. The first author worked with two teams for 5 weeks in person and attended meetings and sessions online for a total of three months before and after the in-person period. The teams we collaborated with work on data driven smart city projects that provide services for both civil servants and citizens. The teams focused on small scale implementations of technical ideas, building prototypes, proof of concept applications, and "first versions" of smart-city technologies. These are then used by decision-makers to discern how ideas could be turned into reality, what kind of opportunities and risks that would bring, and what kind of technical, architectural and user requirements would be necessary to make it work. In this way, both teams provide research, ideation, and prototyping, as well as technical development of smaller scale projects. Following the release of ChatGPT, AI has received increasing attention within the municipality and technical teams working with ML technologies are facing lots of interest from other departments. The team under study in particular considered themselves responsible for showcasing what digital technologies and AI can do for the city.

The first author leveraged their prior expertise as a designer to collaborate with the teams as an embedded researcher on one of their exploratory projects - the development of a routeplanner for people with mobility impairments, that aims to create personalised and targeted routes depending on accessibility needs. The tasks during this project included user research with wheelchair users, design sketches, development of technical requirements, data infrastructure and technical prototypes for route testing.

This project only provided the background for the first author to be integrated into the team. As such, research focused on team tasks and operations more generally during the experience of using ethics tools for several different projects. The first author was able to observe how the team members talk about and perceive their work and its ethical dimensions, as well as how they integrate the practical tasks into their everyday work, negotiate required resources and requirements, and schedule, scope and define resulting work tasks. This allowed the first author to experience how each team defines itself in relation to the rest of the organization, how they see their internal and external responsibilities, and their team purpose. In addition to mapping the logistical challenges, we paid attention to the feelings and emotional layers emerging through reports, team communications and actions. The observed work projects provide illustrative examples for how the affective experience around ethical decision making unfolds. In our analysis, we focus on the positive and negative emotions connected to the processes, rather than the outcomes of the processes by themselves. We therefore focus in this paper on the team that continuously worked on integrating the toolkit into their work practices in order to demonstrate the challenges for an engaged and invested team of practitioners.

The teams worked on up to 3 projects at the same time, thus allowing the first author to observe and learn about the daily operations across multiple project contexts. In addition to participating

in everyday work life as a temporary part of a project team, the first author conducted interviews with all team members, attended regular team sessions such as stand ups, sprint planning, team retrospectives and task refinement sessions, and joined additional team sessions scheduled as needed. Interviews and meetings covered a broad range of past and ongoing projects in the City. The first author also attended joint meetings between the teams and other stakeholders, and interviewed other employees of the city, such as managers or close collaborators. In total, we conducted 13 formal interviews, attended 30 team meetings, and 15 other non-team specific meeting sessions. Interviews were recorded, and meetings were documented through extensive note taking. Throughout the stay, a diary was kept for daily note taking and reflections to document daily observations.

The research team maintained a weekly correspondence where the first author sent regular ethnographic reports to the rest of the team during the fieldwork, engaging in joint discussion and reflection over email to facilitate ethnographic theorization [14]. After the fieldwork, the interviews were transcribed and open coded alongside ethnographic notes, followed by thematic analysis by the first author. Emotions and affective experiences crystallized as recurring and relevant to the teams experience of ethics interventions. During the analysis phase, the research team recognized signs of moral stress in the data and used it as a sensitizing concept to explore the language used to express emotional experiences and relational practices. We then synthesized the findings into their presented format. Study participants gave comments on early drafts of this paper to ensure consistency and correct understanding of city operations.

#### 4.1 The Team

When the first author began fieldwork, the team under study had already run a training workshop with Tanya for the route-planner project, and had been familiar with the Manifesto for about 6 months. As an outcome of the training workshop, the team had selected one of their members to take the role of the *Ethics Owner* to guide their engagement with the manifesto going forward. To provide context for our findings, we first introduce the 5 core members of the team:

- **Sarah** is an AI researcher and has been with the team for 5 years. She has taken on the team lead position in the last year.
- **Paul** is an AI specialist and has joined the team 3 years ago. He has always been interested in inclusion and diversity, pushing for more accessibility projects.
- **Nadine** has been with the team as data scientist for 4 years, and has taken on the role as the team Ethics Owner.
- **Eli** is a junior data analyst. She was invited to join the team a little over a year ago by Sarah, after doing her bachelor thesis project at the municipality with some people from the team.
- **Johan** is the project manager in the team. He has worked at the municipality as a freelancer for a few years and has joined the team officially about a year ago.

After the training workshop with Tanya, the team was prepared to run their own workshop sessions internally. Nadine has copied

the templates for the workshop onto a collaborative online canvas. For each new project, she copies a new template next to the existing ones, so that all of the documentation for previous workshops is in one place and visible during future sessions. Before our visit, the team had run one more workshop by themselves for an ongoing project. During fieldwork, the first author attended one follow up workshop for the route-planner project, and two workshops for newly started or planned projects.

### 5 Findings

Researchers have argued that for ethics interventions to be successful there must be a combination of tools and practices developed with existing work structures and practical requirements in mind [57] as well as emotionally invested and open-minded practitioners willing to engage and reflect [32, 60]. The City has put a lot of effort into creating the requisite tools. As the workshop becomes slowly adopted across the organisation, Tanya reports on the experiences the participants face:

*“[The Workshops] create a space where people that maybe have doubts, somewhere in the back of their minds... it helps you bring out some of the uneasiness, maybe something feels off but doesn't really have a space, like in what meeting are you going to address this lingering tiny thing in your mind? And the workshop was really designed to give everybody a voice to put that on the table. Which can be hard, or tricky. But people do get to interesting insights, I think. Quite often you find that there's tensions in the team, that people do have all kinds of issues with what they're doing, they have their doubts, and then that all comes up during the session.”*

As Tanya mentions, the workshops give the teams the space to surface lingering ethical concerns and to sharpen their moral awareness around the work challenges they face. The training workshops offer an introduction to the general ethics initiative across the organisation, but they are only the kick off for an expected ongoing process for the teams. Teams often leave the workshop with more questions and open disagreements than they were aware of before. In order to settle these, they need to continue to develop their ethical sensitivity, internal vocabulary, and communications around the ethical implications of their projects. It is at this point, where overwhelmed teams might abandon the process and stop engaging with the toolkit moving forward. The team under study however was ethically engaged and committed to continue figuring out new questions and emerging doubts, as evident from their own efforts in repeating the workshops for new projects and their openness to our fieldwork engagement. Multiple team members indicated that they were interested in hearing an external perspective on their efforts, because on some level they felt that perhaps they were doing something wrong in their efforts with the Toolkit. During the very first week, Paul states:

*“I am really interested in hearing what you think after spending time with us! You can tell us what we are doing wrong!”*. Johan, the project manager, is diligent about making sure that we report back to them after the visit, so that they can learn from our insights and observations, and Sarah concurs: *“I mean you saw*

*some of our difficulties we have... It would be great to hear what we can improve. I am sure there is so much we are not doing great at the moment."*

The team's reactions to our visit show that they are committed to ensuring that their work results in ethical outcomes (do the right thing) and that they use the provided materials well to achieve those (do the thing right). Throughout the visit, the team expressed eagerness and anxiety, reaching out to an external expert to be helped and reassured, to *tell them what they are doing wrong*. In this way, they were giving voice to the moral stress they experienced due to the newly developed ethical sensitivity, which expressed itself as an underlying frustration and ongoing struggle to find guidance and avoid uncertainty [55]. As will become visible in the following sections, the conversations where the team reflected upon and questioned their work practices are noticeably laced through with affective language, making visible the moral stress that accompanies moral awareness and is coupled to the ethical tensions being discussed through the ethics workshops. These simmering emotional responses to the ethics efforts caught our attention. The team embraced the questions and ethical implications raised by the Toolkit, and did their best to respond to them. Literature suggests that moral stress is an important concern when developing ethical sensitivity in organisational contexts, especially where such interventions are expected to deliver satisfying results [73]. In the following section, we trace the duality of eagerness and insecurity the team displayed across their work processes. We identify where moral stress comes up in different forms, how it is articulated and navigated by the team, and how it affects them beyond the logistical demands of navigating trade-offs.

## 5.1 Relating ethics across organisational structures

Organisational structures play a big role in how production teams perceive their capabilities and power to address ethical implications of their work, when they become aware of them [57, 94, 99]. The team in our study faced a number of challenges, though partially alleviated through the proceduralised nature of the Toolkit, that provided space and a vocabulary to express ethical sensitivity in regards to their work. This organisational effort gives space for other tensions to become visible - tensions and stress that originate from relational friction, but are ultimately experienced individually [74]. Organisational structures, such as management hierarchies and project schedules, are a necessary part of managing technical projects. However, they not only structure work processes, they also structure how people across and beyond an organisation relate to each other - they create relational configurations, within which different facets of moral stress unfold. In order to understand antecedents and expressions of moral stress we attend to the relational configurations that the team navigates with and through stressful emotions. Other researchers [68, 83] have highlighted the impact that relational factors play in how designers and engineers react and deal with moral pressures. Pillai and colleagues [68] for example used the differentiation between micro, meso, and macro layers of closeness. We find a similar importance of distance and closeness within the relational configurations but ground these in organisational structure: relations within the team, relations beyond team

within organization, and engagements with stakeholders beyond the organization.

**5.1.1 Within Team: Owning Ethics.** The team itself is the first space where tech workers explore and potentially confront ethical dilemmas of their work. After the initial training workshop the team itself provides the closest social frame for negotiation, confrontation, and decision making around ethical implications of the projects. As Ethics Owner, a large part of that ongoing process is driven by Nadine, who schedules and sets up ethics workshops for different projects, and takes care of integrating them into the existing processes. One example of this work is her task to transfer ideas from ethics workshops to the team's backlog. Nadine aims to create explicit ties between the principles the team discuss during the workshops and concrete tasks that she can put on the backlog for the next team sprint. She considers this the best way to integrate the responsibility for ethical implications as shared, "normal" work for the entire team that is integrated with other work tasks, rather than treating them as special efforts. She does this by re-framing the ideas of the workshop as ticket items that are collected in the backlog of the team - an ongoing to do list, where the team collects outstanding tasks that will be taking into consideration for planning the future allocation of time and resources. The team works in three week intervals, called sprints, and at the beginning of each sprint, the team visits the backlogs and plans which tasks from the backlog to pick up in the coming three weeks, who should work on it, and what constitutes successful task completion. Nadine explains:

*"the goal is to get the stuff [the ideas from the ethics workshop] we want to do on the backlog.[...] And anything of reasonable size that we work on has to be here otherwise it will not be considered to be picked up. So our only way to get this stuff [she points at the post its with ideas from the workshop] done is to get it in here [she gestures at the backlog]. Yeah, and the only thing I need to think about is.. I need to think a little bit about formulation. And I need to think a little bit about like, what's the first step."*

This process of transferring ideas originating in the ethics workshops into the ongoing work process is the doing of ethics, where Nadine takes the responsibility of turning a discussion into a first step that is actionable and feels familiar enough for the team to act on it. She sees this as the only way to ensure that implications discussed in the workshops get addressed. The Toolkit provides a vocabulary and a frame for how to engage and apply ethical sensitivity. As a follow up the team finds touch points between the other things that are going on and their insights from workshops. Nadine channels a constant stream of translation and integration between reflections, insights, and work tasks through her role by turning ethical inquiry into more mundane tasks that can survive in the standard formats of project management. Without this translation work of the ethics owner, the ideas from the workshops only exist as good intentions, without a concrete tether to the work reality of the team. From a process point of view, this translation work should act as a mitigation strategy, in that it transitions ideas with moral relevance into technical tasks that can be fulfilled without further reflection. In practice however the task of this translation work focuses moral stress on one person to make moral decisions

for the team, and to provide mitigation for the rest of the team members [55].

Nadine is able to perform this translation work because she has both internalised the ethical commitments of the Manifesto, and is an integral part of the team and its culture of collaboration. As Tanya comments, the teams are in charge of making the actual judgements about which costs they are willing to take on to address the implications uncovered in the workshops. To ensure that ideas evolve into tasks that are narrow enough to be acted upon, Nadine needs to apply her personal judgement to what she considers necessary for a task to be performed in accordance with the ethos of the ideas that spawned it. The moral stress emerges exactly at the seam between reflection and actions, where toolkits and processes are placed to enable the transition [46]. For example, one idea from the workshop is to define a second type of user persona, that has more complicated accessibility needs, and that challenges the team to expand what they consider accessible. While documenting this idea as a ticket item in the backlog, she adds "collaboration with a designer" to the task description:

*"That is something I took the freedom to add myself. So I think in general, it's just not nice to have either the person working with the users or the project leader, not included in this kind of large decision. I think it kind of... it will shape the project quite a bit. I would be worried that somebody would forget it so I took the liberty to write it down. [...] And now, what I'm considering is if I want to include in this ticket also like investigating this user persona and find out what is needed or if I want to do that separately... I think I want to do it as one because otherwise things are just moving so slowly."*

In addition to formulating the first step of turning an idea into action, Nadine applies personal judgements to modify and amend task formulations to ensure that they deliver on their intention. Through these judgements she bridges between the team work practices and the intentions of the ethics workshops. In this case she is aware that the definition of the task, who is included in it and who is not, will have a big impact on how the project turns out, and takes individual measures to adjust task instructions accordingly to comply with the context within which they will be executed. This is a sophisticated interpretation and understanding of the influence of relational set up of tasks that she applies, both in regards to the construction of the task, how it will be received by the team and how it will impact the future of the project. This role of the ethics owner is not only an additional work load to her normal tasks, but one that is explicitly morally charged. While she is personally invested in the quality of her work, and cares about ethical implications a lot, Nadine describes the role as stressful:

*"I really don't like the term Ethics Owner. It feels heavy. I am the kind of person when I have responsibility, I take it very seriously. So if you call me Ethics Owner, it is just too much on my shoulder."*

The heaviness that Nadine notes is a by-product of what ends up being an individualized responsibility for morality [55] that she takes on in the role of the Ethics Owner. The ideas build pressure, they construct the *shoulds* that the team feel hanging above their heads.

While Nadine takes on the responsibility of facilitating the translation from the ethics workshops to the backlog, the team supports her efforts and contributes to sharing the responsibilities by being responsive and inquisitive during follow-up meetings, such as sprint reviews and task refinement, where they plan and evaluate their project work together. Overall, this strategy seems successful, and others in the team explain that they appreciate the way Nadine translates the abstract discussions during the workshops into concrete tasks. In some cases the team calls each other out when they plan too optimistically with their work time, or are too loose in their definition of what defines a task as done. So while Nadine carries a deciding role in the constant stream of reflections and project planning, the team has also integrated this role into the social support network they provide each other, and thus provide mitigating structures by dispensing stress through social support [46]:

*"I think we actually depend quite a lot on reviewing each other's work. It's pretty standard for most of the things we do, that somebody watches along and we expect feedback. And yeah, I actually feel this is going pretty well. So I think, just this pointing out that something is not completed can be uncomfortable and people don't always like it. But more generally, like the reviewing of each other's work, I think that's going pretty well and it's kind of well appreciated by my team. [...] And what I do enjoy is that there's a lot of freedom. I think people are kind. They think really deeply about subjects. [...] there's also quite a lot of room now to debate on these things."* Nadine reflects upon the teams interactions.

In this way, the team applies the same care to the additional tasks resulting from the ethics workshops as with any other part of the project, and support each other in navigating the emerging spaces of uncertainty [79]. The moral stress that arises from the ethical sensitivity that Nadine applies during the translation of ideas into tasks is, if not absorbed, to some extent alleviated through an emotionally responsive team culture. The stable social relations of the team provide an emotional buffer and space for moral stress to be addressed and mitigated through shared recognition [46]. Yet throughout the fieldwork team members repeatedly expressed frustration that integrating ethics insights continued to be challenging, convinced that they must be doing something wrong.

*5.1.2 Beyond team, within organisation: The tech hype and fear of missing out.* This section describes the relations of the team as an entity with a specific identity within the organisation, with respect to the internal stakeholders they communicate with. Conflicting obligations within an organisational context can often cause stress and moral stress in particular emerges when moral obligations are in conflict with organisational roles or mandates [74]. The team is embedded in the innovation department of the municipality, part of a broader organization where they collaborate and communicate with a variety of other teams and stakeholders. The release of ChatGPT has led to a steep rise in interest from other sections in the department, who are curious about using AI in their work processes to make their work easier or more efficient. Nadine describes the current position of the team in the organization where

the municipality is trying to find their direction around the use or non-use of AI technologies:

*"One of the challenges I do think is the tech hype, and that there is like, oh, there's this interesting new technology, we have to do something with it. So people are sort of expecting something, but really not clear about what they're expecting. So you're in some weird middle ground where you have on one hand really a lot of freedom, but at the same time, not that much freedom because so many people are watching you."*

The team is under pressure to deliver on the promises of this exciting new technology, that currently feels like an under-explored opportunity for a digitally progressive city. Just as many companies follow the streams of innovation-culture hype, the municipality lives under the pressure to keep up with the technical developments pushed by media. At the same time, the Toolkit workshops add additional demands on the team to be conscious and reflexive about the motivations and intentions behind their technical work:

*"It seems that even though there's all this fear mongering, I think when people want to use AI for their work, they just want the benefits and they don't want to think about it right,"* states Sarah.

One of the requests is an internal version of ChatGPT for the municipality, which upper management wants to develop "so we can put that out there and make informed expert recommendations". However the team keeps running into confusion about the project's goals and intentions. The question of "who are we doing this for and why are we building this?" comes up in multiple conversations, such as the ethics workshop, the backlog refinement, and the sprint planning. During the ethics workshop for another, similar AI project, the question for the motivation to use LLMs comes up again and Paul wonders:

*"Why do we do this? Are we now automating something that was supposed to be human effort anyway? Or, you know - why is this the solution to the problem? [...] Is it just because LLM is a convenient hammer that makes every problem look like a nail?"*

This expression of friction between the push for innovation-driven fast development and considerate, well informed project decisions arises from the fact that the workshops add more demands and pressure on teams, in the context of a culture that is already charged with time pressures to deliver technical innovations at fast speed. The willingness of the team to reflect beyond the technical feasibility of their potential projects and the underlying intentions of their tech ideas confronts them with the limitations of their own roles and responsibilities. As the technical team, they are the experts that the municipality comes to for consultations about feasibility and testing, but the introduction of the Manifesto and the Toolkit also asks of them to follow their ethical sensitivity to an extent that is not mirrored throughout all the layers of the organisation. Organisational structures provide a structured separation of concerns [84], to enable clear delineations of responsibilities. However, the Toolkit and the training of technical teams in ethical sensitivity result in loosening of formal and informal separation of concerns, leading

to the team being confused and unsure about where to draw the line around their responsibilities.

The City has legal teams that are seen as responsible for ethics-related decisions especially in the context of compliance with regulation. Yet the workshops teach the technical teams to recognise ethical implications that go beyond questions of legal concerns, compliance and regulations:

*"The less clear questions are, of course, the more vague ethical feelings that you might have about something where, you know, you could somehow frame it in a way that meets all the official regulations, but still, your gut feeling says that it's not a great idea."* Johan says.

In these situations, where the municipality is still working on finding direction, the team works to balance their responsibilities as civil servants as well as technical experts. This puts the team into an unofficial role as gate keepers, with the implied responsibility to judge which risks are ethically relevant enough to raise for discussion beyond the team and which are not. As informal gate keepers, they feel responsible to consider the ethical consequences of potential projects, without necessarily having the mandate to change them – something generally associated with legal teams, who act as gatekeepers on legislative grounds:

*"In the end, it's going to be a sort of political choice that a manager at some point needs to sign off on."* Johan shares.

The term "political choice" has a rather specific meaning in a municipal context, highlighting tensions between capital P Politics (as in political programs developed by parties) and small p politics (as in implicit value judgements about who and what to prioritise) which civil servants must navigate within formal governance structures [8]. Capital P "Political Choices" can only be made if there is explicit mandate to do so, and such choices get escalated back up to the political decision layer. "Political Choices" are by definition decisions that civil servants cannot make on their own. Tensions arise in differentiating when a dilemma is a managerial (political) choice, and when it is a Political Choice.

While the team is supported through the framing of the tools to open conversations amongst each other and with other teams and management, the identified way forward is grounded in political decisions that are often out of the team's locus of control. The tools give the team the capacity and space to reflect on the ethical dimensions of their work, but only limited power over the consequences of such reflection and mandate to address them [76]. The team acknowledge their limits and limitations and generally aim to operate within the boundaries of what they can deliver. However, in contrast to the internal team dynamics, the structures in the organisation do not provide the relational resources to navigate these limitations in ways that offer the team assurances about their insecurities. The moral stress that results from the dissolving boundaries around separation of concerns [84] triggered by the workshops has nowhere to go, remaining with the team to deal with.

*5.1.3 Beyond organisation: good intentions and the status quo.* The municipality is part of the broader City governance configuration together with citizens, businesses, and other organizations and the

team has to grapple with the tensions that arise between their own projects and intentions and the outcomes for the citizens they are serving.

The Manifesto postulates high-level principles and values in relation to citizens of the City, for example stating that data usage should benefit everybody, that models should be transparent for everybody to understand, and the like. The Toolkit and workshops have taught the team to think beyond the technical aspects of their work and consider the broader social and political impacts their projects might have. But beyond these reflections, the workshops also build pressure to act in accordance with those insights and construct strong ideals for the team to live up to, with a long list of "shoulds" hanging over them. So while the workshops give the team the tools to *decide* on ways forward, the actual discomfort around getting these tasks done often surfaces outside of the workshops, in particular when it comes to actions that relate to stakeholders outside of the organization.

For example, the team identified an ethical dilemma in a computer vision project. In an attempt to gather more information about the accessibility of the city public infrastructure, the team worked on a model to help measure the real width of sidewalks. While sidewalks are planned with a certain width by city planners, this is later compromised by additions of permanent fixtures such as polls, trees or signs, or temporary obstacles such as roadworks, wrongly parked vehicles or trash. The project was intended to enable support for people with mobility impairments who rely on wheelchairs or other mobility assistance mechanisms and require a particular width to manoeuvre. Having developed the model however, the team realised that the tool they created also brought up a number of questions about surveillance. Aside from estimating sidewalk widths, the algorithm could pick up on citizen digressions such as mis-parked bikes and misplaced flower pots, which could be penalized more efficiently without actually helping wheelchair users. While trying to create a system that would provide one group with more useful information about the city infrastructure, the team realized that they unintentionally created a system that could also reinforce a punitive approach in cases irrelevant to their purpose (i.e. where large flowerpots do not in fact intervene with accessibility) or lead to punishment without any resolution (such as creating fines for misparked bikes without providing sufficient space for both bikes and an accessible sidewalk), something the team was not comfortable with supporting. Reflecting on this project in particular, Paul worried:

*"We were so well meaning, so well-intentioned, and afterwards I thought, oh, should we even do this?"*

Of course, just because an algorithm can be used in such a way doesn't mean it will be. The team put extra effort into ensuring that potential for misuse is recognized, so that it could be curbed proactively. However, given that the teams also want to be transparent about the models they develop and potentially share algorithms with other parties, such worries remain potent. As Sarah explained:

*"We, as a team, have the same responsibility [as all civil servants to raise issues and concerns] and maybe even a stronger responsibility when it comes to AI".*

After this project Eli has started to see the infrastructures of the city differently. She now notices bikes that are parked blocking the

sidewalks and other obstacles in the streets that make the overall infrastructure less accessible. She has started to make sure that she does not park her bike in the way of a potential wheelchair path and she moves other bikes when she can, getting frustrated when she realizes that the existing infrastructure does not take all needs into account similarly. During one ethics workshop for the accessible route planner, she discusses with the others how to draw the attention of officials to the issues of lacking infrastructure and challenges for mobility impaired citizens. The group agrees that to match the ideals of the Manifesto, broader changes to infrastructure are needed, more than they can do with an app. On the one hand, the team feels responsible for providing technical support for a more accessible city. Considering the necessary changes beyond the production of an app also becomes important for their agenda. They therefore decide to keep the topic of infrastructural changes in their documentation and add an additional post-it to their idea board with the task of sharing data on obstacles with officials to advocate for more accessible infrastructure. On the other hand, the team needs to be conscious about their own resources in order to get something done, even if it is not living up to the full scope of the ideal:

*"We don't have the capacity to immediately develop something that works for everybody. I think one thing we have to do here is clearly communicate what the scope of this pilot is, and make sure that it is also very clear for the public that we've consciously made this choice and are aware of the fact that we are not helping everybody." Sarah states.*

The next day, Nadine reflects on how building an app that supports people with mobility impairments might even relieve pressure off other departments to design the city more inclusively in the first place, which would be the opposite of what the team would like to achieve:

*"Like, you have discussions and for some things, you think okay, yeah, we could be a bit more transparent, that'd be cool. But this one was really hitting me in the face like, okay, I think we're doing ... are we not really consolidating the situation instead of changing it?"*

Deep reflections on far reaching consequences often spark more questions and frustrations when tech solutions treat symptoms rather than causes. In the case of the accessible route planner, any fundamental change in infrastructure is far outside of the mandate or jurisdiction of the team. The team struggles with the problem that the systems they encounter in their projects are built into an environment with already established structures that pull towards a status quo. This has two consequences for the team: first, these structures are not in line with the values in the project and thus can easily subvert it; second, the team can only make progress from the given starting point, but tends to judge itself from an ideal interpretation of values and responsibilities promoted by the Toolkit. In attempting to live up to the idealistic values the team began to recognize that sometimes tech solutions are just band-aids to bigger issues. All their efforts to improve the ethical scoring of their production process – by making it more inclusive and transparent – might not create the end result they would prefer,

simply because their potential solutions are tools and not systemic changes.

## 6 Discussion

In our study, the overall effort and initiative of the municipality can be considered a success - not only did the technical team we studied adopt the suggested methods and engaged with them continuously, they also used the tools to push for project changes in line with ethical considerations. However, the findings also highlight that a success story of an ethically invested tech teams plays out differently from what is expected by toolkit designers, business managers, and even technical professionals themselves. The team experience of increased ethical sensitivity *feels* very different from what ethics toolkits and workshops might promise, as they encounter new vulnerabilities and insecurities. Vulnerability is endemic to the process of accepting responsibility for ethical decisions [5]. This is not only a matter of distributing mandates within the organisational structures and creating appropriate processes, but also of accounting for what Popova and colleagues term intense somatic experience of discomfort [70]. While recent work by Popova and colleagues [69] notes the importance of emotions in engaging ethical issues, what they describe maps directly on expressions of moral stress and the most typical mitigation strategies of distancing and abstracting described in the moral stress literature [46, 73]. In the following section, we discuss what we see as transferable insights from our findings on the emergence of moral stress for HCI.

### 6.1 Beyond reflection: Engineers are set up to fail and expected not to care

Reflection is often seen as a necessary component of ethics tools and methods [24, 32] that can result in important insights [58]. Approaches such as VSD and other interventions developed within HCI have been successful at broadening perspectives, including other voices, and providing productive opportunities to foster the development of moral awareness and ethical sensitivity [6, 15, 33, 61, 89, 96, 98]. What has been lacking is a discussion of the costs that these approaches create. As ethics workshops and toolkits teach technical teams to reflect and question, they inevitably create more questions than answers and thus destabilize previously held certainties. Rather than finding clarity, feeling more assured in the goodness of their work, and capable of delivering ethically sound projects, the team in our study experienced moral stress from uncertainty and lack of assurance as they identified and negotiated increasingly difficult trade-offs, which manifested as anxiety and frustration [55]. Ethics toolkits construct the values they consider as ongoing sociopolitical practice [99]. They also construct the people using them, in this case technical practitioners. Many ethics interventions not only consider technical practitioners as not yet ethically engaged, and missing ethical sensitivity in order to build technology responsibly, they also consider technical practitioners as disembodied, impartial actors, who can work through the expected process of reflection, identity work, and required action without much emotional reaction to shifts of perspective and awareness of positionality and responsibility. Our findings show that rather than an indicator of shortcomings or failing work processes, discomforts

from considering ethical implications are indicators of internalising a required level of commitment and emotional investment. In other words, when engaging with any effort to train and increase ethical sensitivity, moral stress is always part of the deal, and an indication of growth rather than a failing [74].

The affective experience of moral stress goes counter to the general narratives of what tools, pipelines, and work processes are expected to do in a technology innovation context, where creative friction is welcome as long as it does not impose on productivity [81]. While no ethics interventions claim to provide solutions to ethical dilemmas, they typically offer guidance or support in making decisions, thus implicitly promising to ease the burden and improve outcomes [15, 64]. This approach originates from a culture of solutionist thinking, where problems are always coupled with solutions, even if the problems are inherently unsolvable [22, 80]. When it comes to ethics interventions, such coupling creates expectations that dealing with and addressing ethical dilemmas through the provided exercises is a skill to be trained, becoming easier with time, and resulting in 'better' outcomes.

These expectations fall apart when aspirational values and idealistic resolutions collide with real-life working conditions, and the interventions fail to provide unquestionable paths of action. Practitioners expected to engage in ethical deliberation during their technical work are automatically set up to fail on one front or the other, never able to fully reach the idealistic vision portrayed by value sets, manifestos or other prescriptive ethics materials. While VSD-informed methods such as value dams [61] offer practical and actionable approaches for addressing value tensions, such processes in themselves do not provide a lens to understand the emotional and affective impact their application has on the professionals adopting them. This is where we suggest moral stress as a beneficial lens - in addition to evaluating the ways in which such tools provide approaches for ethical dilemmas, we argue that, since ideal resolution is impossible, we also need to understand and account for the emotional fall-out of this 'failing'. The moral stress of having to adjust personal perception of technical work as ethically productive is rarely acknowledged, leading to harsh experiences, such as "being hit in the face", as Paul puts it, with the realization that well-meant technical interventions might contribute to consolidating existing exclusionary structures. Feelings of vulnerability then are more than "an active ethical stance" as Popova and colleagues argue [70]. Rather, they are endemic to making ethical considerations explicit in technical processes and practices and can manifest as moral stress.

Where moral stress remains unacknowledged and hence unaddressed, ethics interventions - no matter how ostensibly effective and well designed - will struggle to achieve the hoped for changes. To mitigate the costs of emotional labour and moral stress, people may distance themselves from the points they should intentionally engage, undercutting the original intentions of the interventions [98]. According to prior research on moral stress, one common mitigation technique is limiting the view of the broader picture and narrowing the space of responsibility for the affected workers [46]. This is in direct opposition to ethics interventions, which seek to increase ethical reflection and broaden the scope of relevant concern. Recognising the broader picture and the relevance of one's own work within it is important to the ideals of responsible

computing but this can make it harder to see small wins rather than bigger failures. This is evident in the struggles of the team around their potential beneficial impact vs. supporting the status quo. Yet focusing on small wins can be an effective tactic working for broader changes in constraining environments, what Horgan and Dourish describe as aspects of tempered radicalism in technical practice [43]. This is where 'small p' politics and moral stress connect. In order to mitigate moral stress and be productive in 'small p' politics, limited views of impact are productive but in tension with the general ideal of developing broader moral awareness. This leads to the required next step that follows after the implementation of ethics interventions - from developing moral awareness towards managing moral stress as they negotiate the discomfort of ambivalence and disjunctions between the moral ideals and the constrained realities of technical practice in organizational contexts. Integrating such recognition into work practices demands emotional labour of the team members individually and collectively, in order to accommodate a shifting perception of their roles, tasks, and responsibilities.

Here, the capacity to lean on a community and safe social relations, as productive coping mechanisms for difficult feelings stemming from ethical quandaries, becomes crucial [68]. Our findings align with prior research in nursing studies, that states that a socially healthy team culture is a non-negotiable pre-requisite for teams to productively engage in ethics efforts that continue to put them in difficult and uncomfortable situations [46]. Well-functioning teams are not the domain of HCI, and there is a wealth of literature within organisational studies and management studies that address this. However, team relations must be a consideration when designing ethics interventions, because these inevitably rely on a solid relational foundation within teams and organisations, where critical and difficult topics are being surfaced. The Toolkit, as its creators pointed out to us, is designed to create spaces where doubts can be surfaced and openly interrogated. Such "spaces of doubt" are key to the doings of ethics, even as the ethics owner on the team works to translate these into "nodes of certainty" as concrete steps on the backlog [79]. The team we observed in our study had an extremely supportive and kind culture of openness that enabled them to share the emotional costs of uncertainty and doubt when engaging with the ethical implications of their work. They appreciated this space for reflection but also clearly articulated the increased anxieties and frustrations that resulted in such engagements.

## 6.2 Responsibilities of organizations: Stress that has nowhere to go

Ethics toolkits are always a moral comment on the existing practice, putting teams on the spot to make normative decisions and take moral stances. As our findings show, the spaces for reflection created by the Workshops stand in contrast to a culture that pushes for innovation and being ahead of the curve, creating pressure for engaged teams to both take the time to question and explore and to deliver at the same time. Such pressure ensures that teams ultimately have to fall short on some aspects of their work, accruing new types of social and political costs for the teams. The teams must navigate increasing logistical and procedural trade-offs, such

as spending more time and money on a project to accommodate additional work or changed features. In case of project course corrections, such costs can be political, in that the team needs to put extra effort into negotiating and convincing across layers of hierarchy. Project modifications or course corrections are sometimes not possible or feasible for design or development teams not involved in bigger decision making processes [25, 56]. In these cases, teams may need to compromise on their own ethical standards, potentially resulting in additional moral stress [74].

Part of the challenge is that toolkits localise ethics within technical teams, towards the bottom of organisational structures. As a result, where the team in our study clearly increased their moral awareness, this was not happening evenly throughout the organization. Where ethics practices are anchored in individuals and small teams instead of organizations, ethics interventions fall short of impact and put the individuals into difficult and draining positions, when they must navigate the social costs these interventions impose [72]. Widder and Nafus [94] discuss these relational factors as dislocated accountabilities, arguing that the organisational infrastructures necessary for technical development locate social appropriateness of what can be said, by whom, and with which results. In our study, the team was not being wilfully ignored in their concerns, yet they encountered the limitations of their roles in articulating their anxieties, as organizational structures are not built to attend to what Sarah in our study refers to as "vague gut feelings" when trying to estimate costs and utility benefits from technical solutions.

These limitations are inherent to the nature of toolkits - toolkits try to make things accessible, with a low threshold for engagement [99]. They therefore have to target small circles of impact in order to make them actionable and easily adoptable. Expecting toolkits to be localised anywhere beyond the individuals and small teams is opposite to their purpose. Toolkits can create the motivation and necessity for change, but they can not provide guidance for further necessary organisational changes that are required for outcomes such as moral stress to be absorbed into the relational systems of an organisation. Even if reflection work and ethical sensitivity can be developed from the bottom up, the management of the emotional cost needs to be acknowledged and accounted for from the top. Otherwise, when employees are left to deal with moral stress on their own without sufficient power to create meaningful change, teams run the risk of surrendering to the pressures of moral stress, potentially choosing ignorance and resignation as coping mechanisms [19, 46].

Ethics, in other words, is hard emotional labour and must be recognized as such, in order to provide meaningful options for stress mitigation. Ironically, engaging in it fully in this kind of labor requires recognition that this will add challenge and difficult to existing processes [81]. This is in contrast with common business practices seeking to avoid additional questions or uncertainties that rock the boat both within teams and outside of them [47, 81]. It is clear that for ethics interventions to succeed, teams require emotional scaffolding in addition to logistical resources and the design of interventions must take notice of this requirement. Organizations must internalize that ethics considerations are tricky and projects will become more openly complicated when surfacing

ethical tensions. Feminist scholars [27] call for organisational cultures grounded in feminist notions of dynamic response-ability, a reevaluation of care and maintenance practices and shared responsibilities. Yet even in best case scenarios organisational structures can be notoriously difficult to shift. Establishing explicit support structures and strategies for self-care, such as de-briefing sessions that explicitly give space to raise and share the felt experiences and affective impact of ethics work, and giving permission to admit and voice feelings of inadequacy and vulnerability, are some of ways to mitigate and address the emotional labour baked into technical work [5]. We hope that being able to identify moral stress as a core challenge can contribute to efforts aiming to design interventions that can provide the emotional scaffolding to technical teams grappling with ethical implications, and focus organisational attention to support the necessary within-team collaboration and care practices.

## 7 Conclusion

In this paper, we discuss the affective experience of integrating ethics driven interventions in technical work practices. Even when toolkits are designed with logistical necessities in mind and technologists are emotionally invested and motivated to “do the right thing”, ethics interventions require emotional labour and produce moral stress, in the form of vulnerability, discomfort, and other difficult emotions. We argue that a successful and sustainable implementation of ethics initiatives requires accounting for inevitable moral stress [74]. Beyond the guidance on how to “do” ethics, teams require support in learning how to acknowledge and navigate the affective experiences they encounter, and receive reassurance that uncomfortable emotions are not necessarily a sign of wrong conduct, but an inevitable part of the experience [46].

Ethics interventions are often located within technical teams, without any expectations of shifts in organizational practice. While organizations might be willing to support their technical teams by providing extra time and resources for integrating new ethics-related tools and processes, few recognize the necessity for broader organizational change. We show that teams that embrace moral awareness and attempt to influence organizational decisions accordingly often encounter barriers in the form of managerial hierarchies and internal politics. Such barriers can lead to frustration, disengagement and stunted impact of any intervention that aims to develop moral awareness without political reach. Overlooking the inevitability of moral stress can explain why ethics materials meet little adoption or acceptance [32, 57, 60], especially where teams fully embrace the intentions and expectations of ethics initiatives. Working through ethical tensions therefore needs to be a full-scale organizational commitment, where moral stress is recognized as a legitimate part of ethical work through supportive structures. Going beyond typical organizational resources such as project management and communication structures, organizations must provide emotionally supportive scaffolding that acknowledges the extra costs of moral awareness and moral stress [19, 73].

In closing we want to emphasize that this work is not intended as a criticism for the particular organisation we worked with. Our feedback and findings about moral stress have been met with enthusiasm, where the organisation is willing to do the necessary

work that is required in building in ethical reflection into technical practice. In fact, we believe that we were able to initially identify the manifestations of moral stress so clearly precisely because the team we collaborated with felt safe in openly voicing their doubts and concerns. We see further promising directions for this research in deepening the understanding of how ethics initiatives can address the socio-political dimensions attached to organizational practices and hope to motivate other researchers to follow a path that recognizes practitioners as emotional beings with affective needs when grappling with responsibilities, ethical tensions, and uncertain consequences.

## Acknowledgments

We thank the teams that welcomed us to spend time with them and talk about the difficult topics of ethics in innovation practices. This work is part of the DCODE project, funded by the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 955990. This work has also been supported by the BUGGED project funded by the Research Council of Finland decision number: 348391 and the Nokia foundation’s Jorma Ollila Grant. Views and opinions expressed are those of the authors only and do not necessarily reflect those of the funding organisations.

## References

- [1] 33A. [n. d.]. *AI Design Sprint™*. <https://www.33a.ai/ai-design-sprint>
- [2] Safinah Ali, Vishesh Kumar, and Cynthia Breazeal. [n. d.]. AI Audit: A Card Game to Reflect on Everyday AI Systems. 37, 13 ([n. d.]), 15981–15989. doi:10.1609/aaai.v37i13.26897 Number: 13.
- [3] Justin Ames, James Gaskin, and Bradley D. Goronson. 2020. Exploring antecedents and consequences of managerial moral stress. *Business Ethics* (2020).
- [4] ARTEFACT. [n. d.]. *The Tarot Cards Of Tech*. <https://tarotcardsoftech.artefactgroup.com/>
- [5] Madeline Balaam, Rob Comber, Rachel E. Clarke, Charles Windlin, Anna Ståhl, Kristina Höök, and Geraldine Fitzpatrick. 2019. Emotion Work in Experience-Centered Design. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (2019).
- [6] 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*. 421–433.
- [7] Paula Boddington. 2017. *Towards a code of ethics for artificial intelligence*. Springer.
- [8] Kirsten Boehner and Carl DiSalvo. 2016. Data, Design and Civics: An Exploratory Study of Civic Tech. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) (*CHI '16*). Association for Computing Machinery, New York, NY, USA, 2970–2981. doi:10.1145/2858036.2858326
- [9] Karen Boyd. 2022. Designing Up with Value-Sensitive Design: Building a Field Guide for Ethical ML Development. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (Seoul, Republic of Korea) (*FAccT '22*). Association for Computing Machinery, New York, NY, USA, 2069–2082. doi:10.1145/3531146.3534626
- [10] 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 (2021), 1–29.
- [11] Francesca Bria, Malcolm Bain, Guillem Marpons, Javier Usobiaga, Marta Armada, and Ivan Minguenza. 2017. *Barcelona Ethical Digital Standards Policy Toolkit*. <https://www.barcelona.cat/digitalstandards/en/init/0.1/index.html>
- [12] Francesca Bria, Malcolm Bain, Guillem Marpons, Javier Usobiaga, Marta Armada, and Ivan Minguenza. 2017. *Barcelona Manifesto in favour of technological sovereignty and digital rights for cities*. <https://www.barcelona.cat/digitalstandards/manifesto/0.2/>
- [13] Ania Calderon, Dan Taber, and Hong Qu Jeff Wen. [n. d.]. *AI Blindspot: A Discovery Process for preventing, detecting, and mitigating bias in AI systems*. <https://aiblindspot.media.mit.edu/>
- [14] Allaine Cerwonka and Liisa H Malkki. 2008. *Improvising theory: Process and temporality in ethnographic fieldwork*. University of Chicago Press.

- [15] Shruthi Sai Chivukula, Ziqing Li, Anne C Pivonka, Jingning Chen, and Colin M Gray. 2021. Surveying the landscape of ethics-focused design methods. *arXiv preprint arXiv:2102.08909* (2021).
- [16] 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*. 1–13.
- [17] Amsterdam City. 2020. *Amsterdam Tada Manifesto*. <https://tada.city/en/home-en/>
- [18] Amsterdam City. 2023. *Fiawo, Jason Creating An Accessible City and Monitoring Ethics*. <https://amsterdamintelligence.com/posts/creating-an-accessible-city>
- [19] Mary C Corley, Ronald K Elswick, Martha Gorman, and Theresa Clor. 2001. Development and evaluation of a moral distress scale. *Journal of advanced nursing* 33, 2 (2001), 250–256.
- [20] Nicholas Kluge Corrêa, Camila Galvão, James William Santos, Carolina Del Pino, Edson Pontes Pinto, Camila Barbosa, Diogo Massmann, Rodrigo Mambri, Luiza Galvão, Edmund Terem, et al. 2023. Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance. *Patterns* 4, 10 (2023).
- [21] Josh Cows, Thomas King, Mariarosaria Taddeo, and Luciano Floridi. 2019. Designing AI for social good: Seven essential factors. Available at SSRN 3388669 (2019).
- [22] Jay Cunningham, Gabrielle Benabdallah, Daniela Rosner, and Alex Taylor. 2023. On the grounds of solutionism: Ontologies of blackness and HCI. *ACM Transactions on Computer-Human Interaction* 30, 2 (2023), 1–17.
- [23] Wesley Hanwen Deng, Nur Yildirim, Monica Chang, Motahhare Eslami, Kenneth Holstein, and Michael Madaio. [n. d.]. Investigating Practices and Opportunities for Cross-functional Collaboration around AI Fairness in Industry Practice. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (New York, NY, USA, 2023-06-12) (FAccT '23). Association for Computing Machinery, 705–716. doi:10.1145/3593013.3594037
- [24] Richard Devon and Ibo van de Poel. [n. d.]. Design Ethics: The Social Ethics Paradigm. 20, 3 ([n. d.]), 461–469. <https://www.ijee.ie/articles/Vol20-3/IJEE2514.pdf>
- [25] Christian Dindler, Peter Gall Krogh, Kasper Tikær, and Peter Skov Nørregård. 2022. Engagements and articulations of ethics in design practice. *International Journal of Design* 16, 2 (2022), 47–56.
- [26] Dot.Everyone. [n. d.]. *Consequence Scanning – an agile practice for responsible innovators – doteveryone*. <https://doteveryone.org.uk/project/consequence-scanning/>
- [27] Eleanor Drage, Kerry McInerney, and Jude Browne. 2024. Engineers on responsibility: feminist approaches to who's responsible for ethical AI. *Ethics and Information Technology* 26, 1 (2024), 4.
- [28] W Scott Dunbar. 2005. Emotional engagement in professional ethics. *Science and Engineering Ethics* 11 (2005), 535–551.
- [29] Jessica Fjeld, Nele Achten, Hannah Hilligoss, Adam Nagy, and Madhulika Sri Kumar. 2020. Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI. *Berkman Klein Center Research Publication* 2020-1 (2020).
- [30] Luciano Floridi. 2015. *The onlife manifesto: Being human in a hyperconnected era*. Springer nature.
- [31] Luciano Floridi, Josh Cows, Monica Beltrametti, Raja Chatila, Patrice Chazerand, Virginia Dignum, Christoph Luetge, Robert Madelin, Ugo Pagallo, Francesca Rossi, et al. 2018. AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. *Minds and machines* 28 (2018), 689–707.
- [32] Christopher Frauenberger, Marjo Rauhala, and Geraldine Fitzpatrick. 2017. In-action ethics. *Interacting with computers* 29, 2 (2017), 220–236.
- [33] Batya Friedman and David G Hendry. 2019. *Value sensitive design: Shaping technology with moral imagination*. MIT Press.
- [34] Ester Fritsch, Irina Shklovski, and Rachel Douglas-Jones. 2018. Calling for a revolution: An analysis of IoT manifestos. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–13.
- [35] 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*. 1–15.
- [36] Jet Gispén. [n. d.]. *Ethics for Designers*. <https://www.ethicsfordesigners.com>
- [37] Deborah B Gould. 2009. *Moving politics: Emotion and ACT UP's fight against AIDS*. University of Chicago Press.
- [38] 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*. 1–11.
- [39] Colin M Gray, Shruthi Sai Chivukula, Thomas V Carlock, Ziqing Li, and Ja-Nae Duane. 2023. Scaffolding Ethics-Focused Methods for Practice Resonance. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*. 2375–2391.
- [40] Colin M Gray, Ikechukwu Obi, Shruthi Sai Chivukula, Ziqing Li, Thomas Carlock, Matthew Will, Anne C Pivonka, Janna Johns, Brookley Rigsbee, Ambika R Menon, et al. 2022. Practitioner trajectories of engagement with ethics-focused method creation. *arXiv preprint arXiv:2210.03002* (2022).
- [41] Ben Green. [n. d.]. The Contestation of Tech Ethics: A Sociotechnical Approach to Technology Ethics in Practice. 2, 3 ([n. d.]), 209–225. doi:10.23919/JSC.2021.0018 Conference Name: Journal of Social Computing.
- [42] Thilo Hagendorff. 2020. The ethics of AI ethics: An evaluation of guidelines. *Minds and machines* 30, 1 (2020), 99–120.
- [43] Leah Horgan and Paul Dourish. 2018. Ambiguity, Ambivalence, and Activism: Data Organizing Inside the Institution. *Krisis | Journal for Contemporary Philosophy* 38, 1 (2018), 72–84. doi:10.21827/krisis.38.1.37183
- [44] Hyper Island. [n. d.]. *Unintended Consequences*. <https://toolbox.hyperisland.com/unintended-consequences>
- [45] Javier Camacho Ibáñez and Mónica Villas Olmeda. 2022. Operationalising AI ethics: how are companies bridging the gap between practice and principles? An exploratory study. *Ai & Society* 37, 4 (2022), 1663–1687.
- [46] Dilini I Imbulana, Peter G Davis, and Trisha M Prentice. 2021. Interventions to reduce moral distress in clinicians working in intensive care: A systematic review. *Intensive and Critical Care Nursing* 66 (2021), 103092.
- [47] Lilly Irani. 2019. *Chasing innovation: Making entrepreneurial citizens in modern India*. Princeton University Press.
- [48] Anna Jobin, Marcello Lenca, and Effy Vayena. 2019. The global landscape of AI ethics guidelines. *Nature machine intelligence* 1, 9 (2019), 389–399.
- [49] Jennifer Jordan. 2009. A social cognition framework for examining moral awareness in managers and academics. *Journal of Business Ethics* 84 (2009), 237–258.
- [50] Eda Kranakis. 2004. Fixing the blame: Organizational culture and the Quebec Bridge collapse. *Technology and culture* 45, 3 (2004), 487–518.
- [51] Q Vera Liao, Daniel Gruen, and Sarah Miller. 2020. Questioning the AI: informing design practices for explainable AI user experiences. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–15.
- [52] Sharon Lindberg, Petter Karlström, and Sirkku Männikkö Barbutiu. 2023. Cultivating ethics with professional designers. (2023).
- [53] Sharon Lindberg, Chiara Rossitto, Ola Knutsson, Petter Karlström, and Sirkku Männikkö Barbutiu. 2024. Doing Good Business? Design Leaders' Perspectives on Ethics in Design. *Proceedings of the ACM on Human-Computer Interaction* 8, GROUP (2024), 1–22.
- [54] Janet Lyon. 1999. *Manifestoes: Provocations of the modern*. Cornell University Press.
- [55] K. Lütznén and B. E. Kvist. 2012. Moral distress: a comparative analysis of theoretical understandings and inter-related concepts. *HEC forum : an interdisciplinary journal on hospitals' ethical and legal issues* (2012). doi:10.1007/s10730-012-9178-9
- [56] Shuhao Ma, John Zimmerman, Sarah E Fox, Valentina Nisi, and Nuno Jardim Nunes. 2024. "My Sense of Morality Leads to My Suffering, Battling, and Arguing": The Role of Platform Designers in (Un)Deciding Gig Worker Issues. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference* (Copenhagen, Denmark) (DIS '24). Association for Computing Machinery, New York, NY, USA, 3501–3514. doi:10.1145/3643834.3660713
- [57] Michael A. Madaio, Luke Stark, Jennifer Wortman Vaughan, and Hanna Wallach. [n. d.]. 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* (New York, NY, USA, 2020-04-23) (CHI '20). Association for Computing Machinery, 1–14. doi:10.1145/3313831.3376445
- [58] Annette N Markham. 2020. Taking data literacy to the streets: Critical pedagogy in the public sphere. *Qualitative Inquiry* 26, 2 (2020), 227–237.
- [59] Andrew McNamara, Justin Smith, and Emerson Murphy-Hill. 2018. Does ACM's code of ethics change ethical decision making in software development?. In *Proceedings of the 2018 26th ACM joint meeting on european software engineering conference and symposium on the foundations of software engineering*. 729–733.
- [60] Jacob Metcalf, Emanuel Moss, et al. 2019. Owning ethics: Corporate logics, silicon valley, and the institutionalization of ethics. *Social Research: An International Quarterly* 86, 2 (2019), 449–476.
- [61] Jessica K. Miller, Batya Friedman, Gavin Jancke, and Brian Gill. 2007. Value tensions in design: the value sensitive design, development, and appropriation of a corporation's groupware system. In *Proceedings of the 2007 ACM International Conference on Supporting Group Work* (Sanibel Island, Florida, USA) (GROUP '07). Association for Computing Machinery, New York, NY, USA, 281–290. doi:10.1145/1316624.1316668
- [62] Brent Mittelstadt. 2019. Principles alone cannot guarantee ethical AI. *Nature machine intelligence* 1, 11 (2019), 501–507.
- [63] Georgina Morley, Jonathan Ives, Caroline Bradbury-Jones, and Fiona Irvine. 2019. What is 'moral distress'? A narrative synthesis of the literature. *Nursing ethics* 26, 3 (2019), 646–662.
- [64] 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 (2020), 2141–2168.

- [65] Luke Munn. 2023. The uselessness of AI ethics. *AI and Ethics* 3, 3 (2023), 869–877.
- [66] Sam Nutt and Jay Saggarr. 2020. *LOTI London Data Ethics Service*. <https://loti.london/resources/london-data-ethics-service/>
- [67] OPEN. [n. d.]. *The Data Ethics Canvas*. <https://theodi.org/insights/tools/the-data-ethics-canvas-2021/>
- [68] Ajit G Pillai, Thida Sachathap, and Naseem Ahmadpour. 2022. Exploring the experience of ethical tensions and the role of community in UX practice. In *Nordic Human-Computer Interaction Conference*. 1–13.
- [69] Kristina Popova, Claudia Figueras, Kristina Höök, and Airi Lampinen. 2024. Who Should Act? Distancing and Vulnerability in Technology Practitioners' Accounts of Ethical Responsibility. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (2024), 1–27.
- [70] Kristina Popova, Rachael Garrett, Claudia Núñez-Pacheco, Airi Lampinen, and Kristina Höök. 2022. Vulnerability as an ethical stance in some design processes. In *Proceedings of the 2022 CHI conference on human factors in computing systems*. 1–13.
- [71] Alison B Powell, Funda Ustek-Spilda, Sebastián Lehuédé, and Irina Shklovski. 2022. Addressing ethical gaps in 'Technology for Good': Foregrounding care and capabilities. *Big Data & Society* 9, 2 (2022), 20539517221113774.
- [72] Inioluwa Deborah Raji, Morgan Klaus Scheuerman, and Razvan Amironesei. 2021. You can't sit with us: Exclusionary pedagogy in ai ethics education. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*. 515–525.
- [73] Scott J Reynolds and Jared A Miller. 2015. The recognition of moral issues: Moral awareness, moral sensitivity and moral attentiveness. *Current Opinion in Psychology* 6 (2015), 114–117.
- [74] Scott J Reynolds, Bradley P Owens, and Alex L Rubenstein. 2012. Moral stress: Considering the nature and effects of managerial moral uncertainty. *Journal of Business Ethics* 106 (2012), 491–502.
- [75] Malak Sadek, Emma Kallina, Thomas Bohné, Céline Mougénou, Rafael A Calvo, and Stephen Cave. 2024. Challenges of responsible AI in practice: scoping review and recommended actions. *AI & SOCIETY* (2024), 1–17.
- [76] John S Seberger, Marissel Llavore, Nicholas Nye Wyant, Irina Shklovski, and Sameer Patil. 2021. Empowering resignation: There's an app for that. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–18.
- [77] Katie Shilton. 2013. Values levers: Building ethics into design. *Science, Technology, & Human Values* 38, 3 (2013), 374–397.
- [78] Katie Shilton. 2018. Engaging values despite neutrality: Challenges and approaches to values reflection during the design of internet infrastructure. *Science, Technology, & Human Values* 43, 2 (2018), 247–269.
- [79] Irina Shklovski and Carolina Némethy. 2023. Nodes of certainty and spaces for doubt in AI ethics for engineers. *Information, Communication & Society* 26, 1 (2023), 37–53.
- [80] Miguel Sicart and Irina Shklovski. 2020. 'Pataphysical Software: (Ridiculous) Technological Solutions for Imaginary Problems. In *Proceedings of the 2020 ACM designing interactive systems conference*. 1859–1871.
- [81] David Stark and Monique Girard. 2009. Creative friction in a new-media start-up. *The sense of dissonance: Accounts of worth in economic life* (2009), 81–117.
- [82] Luke Stark. 2021. Apologos: A lightweight design method for sociotechnical inquiry. *Journal of Social Computing* 2, 4 (2021), 297–308.
- [83] Norman Makoto Su, Amanda Lazar, and Lilly Irani. 2021. Critical affects: Tech work emotions amidst the techlash. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–27.
- [84] Hariharan Subramonyam, Jane Im, Colleen Seifert, and Eytan Adar. 2022. Solving Separation-of-Concerns Problems in Collaborative Design of Human-AI Systems through Leaky Abstractions. 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 481, 21 pages. doi:10.1145/3491102.3517537
- [85] Tangible. [n. d.]. *Ethical Compass*. <https://tangible.is/en/ethical-compass>
- [86] Linnet Taylor and Lina Dencik. 2020. Constructing commercial data ethics. *Technology and Regulation* 2020 (2020), 1–10.
- [87] Nanna Bonde Thylstrup. 2019. *The politics of mass digitization*. MIT Press.
- [88] Eva Tkautz. [n. d.]. *Cards for Humanity: an inclusive design tool*. <https://medium.com/ideas-by-idean/get-out-of-your-bubble-how-we-designed-cards-for-humanity-an-inclusive-design-tool-e486e20b9626>
- [89] Steven Umbrello and Ibo van de Poel. [n. d.]. Mapping value sensitive design onto AI for social good principles. 1, 3 ([n. d.]), 283–296. doi:10.1007/s43681-021-00038-3
- [90] Ville Vakkuri, Kai-Kristian Kemell, Marianna Jantunen, Erika Halme, and Pekka Abrahamsson. 2021. ECCOLA — A method for implementing ethically aligned AI systems. *Journal of Systems and Software* 182 (2021), 111067. doi:10.1016/j.jss.2021.111067
- [91] Loïs Vanhée and Melania Borit. 2022. Ethical by designer-how to grow ethical designers of artificial intelligence. *Journal of Artificial Intelligence Research* 73 (2022), 619–631.
- [92] Qiaosi Wang, Michael Madaio, Shaun Kane, Shivani Kapania, Michael Terry, and Lauren Wilcox. 2023. Designing responsible ai: Adaptations of ux practice to meet responsible ai challenges. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–16.
- [93] Kathryn Weaver. 2007. Ethical sensitivity: state of knowledge and needs for further research. *Nursing ethics* 14, 2 (2007), 141–155.
- [94] David Gray Widder and Dawn Nafus. 2023. Dislocated accountabilities in the "AI supply chain": Modularity and developers' notions of responsibility. *Big Data & Society* 10, 1 (2023), 20539517231177620.
- [95] 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*. 467–479.
- [96] Till Winkler and Sarah Spiekermann. 2021. Twenty years of value sensitive design: a review of methodological practices in VSD projects. *Ethics and Information Technology* 23 (2021), 17–21.
- [97] Richmond Y Wong. 2021. Tactics of soft resistance in user experience professionals' values work. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–28.
- [98] Richmond Y Wong, Karen Boyd, Jake Metcalf, and Katie Shilton. 2020. Beyond checklist approaches to ethics in design. In *Companion Publication of the 2020 Conference on Computer Supported Cooperative Work and Social Computing*. 511–517.
- [99] Richmond Y Wong, Michael A Madaio, and Nick Merrill. 2023. Seeing like a toolkit: How toolkits envision the work of AI ethics. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–27.
- [100] Nur Yildirim, Mahima Pushkarna, Nitesh Goyal, Martin Wattenberg, and Fernanda Viégas. 2023. Investigating how practitioners use human-ai guidelines: A case study on the people+ ai guidebook. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–13.
- [101] Katherine Zou. [n. d.]. *Design Ethically Toolkit*. <https://www.designethically.com>