Designing with Fallibility: Examining the Knowledge Politics of Agency, Methods, and Motivations in Robot Failure Research

Dmitry Muravyov
Ethics & Philosophy of Technology Section
TU Delft
Delft, Netherlands
d.v.muravev@tudelft.nl

Nazli Cila
Human-Centered Design
TU Delft
Delft, Netherlands
n.cila@tudelft.nl

Abstract

A line of research in HCI and HRI has started to consider robot failures, errors, and breakdowns not as problems to be eliminated, but as opportunities to inform and enrich design. This shift has led to growing interest in how robotic fallibility affects user trust, interaction quality, and system acceptance. In this paper, we inquire into what it means to design with fallibility. Drawing on feminist technoscience, we examine how current approaches frame the roles of designers and users (agency), how research methods shape the phenomena they study (performativity), and how underlying research goals carry ethical and epistemological implications (motivation). In recognizing robotic fallibility as a sociotechnical phenomenon and design research as a world-making practice, we provide design considerations that promote more reflexive, inclusive, and politically aware engagements with (robot) failure in HRI and HCI.

CCS Concepts

• **Human-centered computing**; • **Human computer interaction (HCI)**; • **HCI theory, concepts and models**;

Keywords

Robot fallibility, Feminist technoscience, Human-Robot Interaction, Design Epistemology, Politics of research, Robot Failure, Robot Error

ACM Reference Format:
Dmitry Muravyov and Nazli Cila. 2026. Designing with Fallibility: Examining the Knowledge Politics of Agency, Methods, and Motivations in Robot Failure Research. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3772318.3791110

1 Introduction

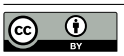
In recent years, videos of robots stumbling, collapsing, or struggling to complete simple tasks, such as opening a door or navigating stairs, have circulated widely online, often to viral acclaim. Shared across social media and news outlets, these clips are typically consumed for their entertainment value, offering viewers a spectacle of technological clumsiness. They play into a broader cultural narrative of robots as awkward, almost-human machines whose missteps are both humorous and strangely reassuring.

Yet, these failures do more than amuse; they expose underlying design values and invite critical reflection on how robotic performance is judged, interpreted, and culturally framed. The attention to robotic breakdown is not limited to popular media; it is echoed in the fields of robotics, Human-Robot Interaction (HRI), and Human-Computer Interaction (HCI), where researchers are acutely aware that robots can err, malfunction, or fail in other ways. Scholars have increasingly turned their attention to what we refer to as *robotic fallibility*, i.e., studying robot failures, errors, and breakdowns, including why and how they emerge, how they are perceived, and how they inform design (e.g. [39, 43, 63]). This research has yielded valuable insights into user experience, dynamics of trust, and system robustness. Centrally, within much of this work, fallibility is embraced so that it can be further mitigated—something to *design with.* It is considered a moment of "epistemic opportunity", typically positioned as a means of improving system performance or interaction quality.

Yet this framing raises key questions. Who defines what counts as failure? How do power, context, and method shape what is learned from these moments? What are the broader implications of treating robotic fallibility as design material? In other words, what remains underexplored is how the concept of "learning from failure" in robotics is situated socially, ethically, and politically. As Soden et al. [97] caution, seeking design knowledge from learning from past failures, unintended consequences, and disruptions risks reducing people and places to "data" and becoming an exercise in objectification and instrumentalization, if not undertaken with care. Current accounts of learning often privilege specific subjects, perspectives, and goals—frequently aligned with institutional priorities or technological efficiency—while obscuring others.

Building on HCI scholarship that examines breakdowns as moments of critical insight into sociotechnical relations (see Section 2), this paper seeks to explore the implications of viewing robotic fallibility as a material for design inquiry. We ask: What does it mean to design *with* robotic fallibility? What can be known, how is knowledge produced, and why might such knowing matter? By addressing these questions, our goal is to open space for more reflective engagements with failure—ones that foreground positionality, power, and the politics of design knowledge.

To explore these questions, we draw on feminist technoscience scholarship as a critical lens for analysis. This family of approaches highlights the situatedness of knowledge, the differentiated positions of knowing subjects, and the intimate connections between



CHI '26, Barcelona, Spain

ACM ISBN 979-8-4007-2278-3/2026/04
https://doi.org/10.1145/3772318.3791110

the research space and the world at large [24, 33, 54]. This perspective entails acknowledging that failures, breakdowns, and errors are not only a technical phenomenon: feminist technoscience inquiry blurs the lines between technical and social, showing the multitude of ways in which technical, engineering, and design solutions have always been social decisions. By reading across existing studies of robotic fallibility through this lens, we aim to surface epistemological, social, political, and ethical implications that are often left implicit in this scholarship.

Rather than provide an exhaustive literature review on robotic fallibility (for which see [43, 63]), our goal is to introduce a theoretical intervention, one that brings feminist technoscience into conversation with robotics and HRI scholarship. We use this framework to examine how failure is constructed, studied, and designed. Therefore, our first intended contribution with this paper is to frame robotic fallibility as a contested and socially embedded category, shaped by researcher positionality, methodological choices, and institutional incentives. Centrally, this contribution thus argues that the robotic fallibility is not only a learning opportunity and a material for design inquiry, but also something that is socially ambiguous, methodologically performed, and leveraged for specific objectives. In doing so, we also provide design considerations to complement current approaches to failure in HCI, HRI, and related fields, or inspire new ones grounded in care, reflexivity, and pluralism.

The paper is structured as follows. In Section 2, we provide a short overview of the robotic fallibility scholarship, relating it to broader discussions on fallibility in HRI and HCI. In Section 3, we introduce key ideas from feminist technoscience to highlight the questions about the subject of knowledge, the performativity of the methods, and the purposes of research. Section 4 connects these conceptual considerations with the discussion on the current research on robotic fallibility. Section 5 provides design recommendations and considerations based on our interpretive analysis.

2 Failure in HCI and HRI

Failure is hardly unfamiliar terrain in HCI and HRI. In HCI, one influential critique targets the rhetoric of "fail fast, fail forward", a central tenet of agile development and certain strands of design thinking. Here, error is framed as a driver for rapid iteration in the service of efficiency and product refinement. Scholars such as Shorey et al. [96], Berger et al. [9], and Sheahan et al. [95] argue that such fail-centered processes obscure the value of frictions and breakdowns, and what these moments reveal about the worlds we inhabit and help construct. From this perspective, breakdowns are not merely steps toward a smoother end product but epistemic events that surface tacit assumptions, expose embedded values, and invite broader reflection on the sociotechnical systems of which they are part [18]. This orientation is echoed in diverse HCI work that treats friction and failure as generative: from rethinking domestic technology breakdowns [114] to designing purposeful failures that surface creative or ethical possibilities [10, 41, 42, 46]. Collectively, these works recast breakdowns as resources—generative openings for reimagining relations among people, things, and design itself.

While much of the HCI literature focuses on failure as an interactional event, a smaller but equally valuable body of scholarship attends to the broader conditions that give rise to failure—that is, the environments, practices, and institutional pressures under which failure unfolds. Vines [117] interrogates how foundational disciplinary assumptions in HCI, such as the mechanistic and computational views of cognition, shape what is labeled as user failure. Gaver et al. [28] candidly recount the failure of their Home Health Monitor, foregrounding the difficulty of defining and evaluating failure in open-ended, interpretive systems, and demonstrating how the pursuit of "design for research" frames how failure is recognized. Ståhl and Tholander [101] reflect on how CHI's peer review conventions pushed their forest-walk design project into a linear, instrumentalized narrative that stripped away its openness and richness, revealing how publishing norms delimit what counts as legitimate knowledge. Finally, Howell et al. [44] use retrospective trioethnography to revisit prior "successful" projects, surfacing invisible labor, mismatched designer–participant imaginaries, participation burdens, and institutional pressures. Together, these works foregrounded the material, organizational, and relational circumstances that shape what counts as failure, and for whom. Their accounts reposition failure not as an isolated event to be fixed, but as a situated outcome of the professional, methodological, and political economies in which design research is embedded.

In HRI, such reflexive critique is rarer. Where failure is studied, it is often instrumentalized to improve performance, usability, or trust [43]. A significant portion of this work explores user responses and recovery strategies. Liu et al. [63] map common cognitive, emotional, and behavioral user responses to robot errors and group recovery strategies, categorizing them into robot-initiated and human-led interventions. Mirnig et al. [73] find that deliberately introducing errors can increase robot likability, as users respond with social signals such as gaze shifts and laughter. Recovery strategies often draw from human-inspired behaviors, such as justifying, promising improvement, or denying mistakes [83]; tailoring responses to users' relational or utilitarian orientations [58]; apologizing [15]; and expressing embarrassment [76]. These strategies have been shown to mitigate the impacts of breakdowns and help repair trust. A smaller strand of HRI research addresses failure in terms of social norms. Lawrence et al. [56] show how conformity to or violation of norms shapes trust, acceptance, and comfort, highlighting the cultural variability of failure perception. Serholt et al. [94] examine how children handle interaction breakdowns with a classroom robot tutee in ways that reveal unspoken rules and expectations in educational settings.

Only a small subset of HRI scholarship engages with failure as a socially and institutionally situated phenomenon. Kamino et al. [48] highlight the sociotechnical infrastructures that sustain long-term human–robot relationships, showing how breakdowns and repairs make visible the cultural norms, emotional labor, and interpersonal relationships that are usually hidden when systems run smoothly. Pelikan et al. [77] show how "wizards" managing robots in a public deployment collaboratively improvise recovery strategies, such as bumping one robot into another to stage autonomy and manage breakdowns in situ, highlighting the invisible human labor behind seemingly autonomous performance. Sætra [88] introduces the concept of strategic robot failure—deliberately

engineering apparent mistakes to achieve hidden objectives—while warning of its ethical implications. Much like self-reflective accounts in HCI, these works reposition failure as a relational and experiential event embedded in the material, organizational, and political conditions of human–robot interaction.

This paper builds on and extends this tradition, aligning with the emerging field of *critical HRI*, which calls for more reflective, ethical, and contextually grounded approaches to the design and study of robots [65, 93]. Critical HRI challenges the dominant framing of robots as primarily engineering solutions to social problems—an orientation that risks obscuring the value-laden assumptions, power dynamics, and unintended consequences embedded in their development and deployment [68]. While recent scholarship has begun to question these assumptions, studies that explicitly examine how research methods, design processes, and researcher positionality shape findings remain scarce, with a few notable exceptions [29, 65, 122, 123].

By approaching robotic fallibilities not merely as scientifically legible, tractable, and measurable states of technical artifacts but also as socially and institutionally situated events, practices, and experiences, this paper responds to calls for research that interrogates underlying design values and makes visible the conditions that shape human–robot interaction.

3 Feminist technoscience and the politics of knowing

To better understand the knowledge produced through studies of robotic fallibility, we draw on insights from feminist technoscience, an interdisciplinary field that examines science and technology through a feminist lens [106, 118]. Central to this tradition is the recognition that knowledge production is never realized from a supposedly detached position, but instead is always partial, embodied, and situated in a certain way, shaped by the positionality of the knowers, institutional structures, and sociopolitical contexts [19, 37, 38, 50, 67]. Feminist technoscience offers conceptual tools for interrogating how technoscientific knowledge is constructed, legitimized, and mobilized. In the context of robotics, it not only reframes what counts as knowledge but also prompts critical questions about the *conditions* and *consequences* of knowing [61, 62, 87, 112]. The insights of feminist technoscience scholarship have also influenced new research directions in the HCI and HRI communities [5, 6, 91, 123].

In this section, we foreground three questions that frame our inquiry: (1) Who and in what way is positioned as the subject of knowledge? (2) What roles do research methods play beyond controlled settings, and (3) What is the purpose of knowledge? These questions draw from a long and intellectually productive lineage of feminist technoscience scholarship, allowing us to examine the often-overlooked social, political, and disciplinary conditions under which knowledge about robot failure is produced. Next, we introduce these questions, as they have been staged in feminist technoscience literature. Their application to studies of robotic fallibility will be examined in Section 4.

3.1 Who is positioned as the subject of knowledge?

This question is a foundational concern in feminist technoscience. Feminist scholars, such as Haraway (1988), Harding (2015), and Longino (1990), have challenged the notion of a disembodied and objective knower [37, 38, 67]. They accentuate, albeit in occasionally contrasting terms, that knowing is always connected to a particular form of social situatedness or positionality that a subject inhabits.

This means that some knowledge claims may be underappreciated. This can happen not necessarily because such claims themselves are less epistemically valuable, but rather because the differentiated social positioning can determine the epistemic value of actors' claims. To understand why specific claims are epistemically (dis)valued, feminist approaches suggest paying attention to how positions of different knowers are constructed in the scholarship on a particular topic.

In AI and robotics, one canonical example is Alison Adam's [1] research on the early AI projects (Cyc and Soar in the late 1980s). These projects claimed to model the general mechanisms of human problem-solving. At the same time, despite its universal ambitions, as Adam explains, the notion of the knower was constructed in a specifically constrained manner, as the projects relied on a very particular sample of technically educated, young, male college students in the US. Thus, the scientific aspirations to obtain knowledge about general principles of human cognitive processes, which have later become crucial for some AI developments, are, in fact, shaped by a vision of a knower that is, amongst other things, a very gendered one.

For our purposes, this question prompts a critical reading of research on robotic fallibility: Who are the knowing subjects that emerge in the literature on robotic fallibility? What are these subjects envisioned to know before, during, and after the robotic breakdown happens? What is the relationship between these knowing subjects? What are these knowers assumed to do when using the pre-given or obtained knowledge? By posing these questions, we aim to surface the often-unspoken assumptions about subjectivity and agency that organize the field. This, in turn, opens space to consider whose perspectives are centered, whose are marginalized, and how these dynamics shape both the design and interpretation of failure in robotics.

3.2 What roles do research methods play beyond scientific settings?

Feminist technoscience recognizes that knowledge is always produced in specific contexts, and it is from these contexts that aspirations to universal knowledge become possible. Rather than treating experimental constraints as limitations to generalizability, feminist scholars argue that these conditions actively shape what counts as knowledge and which realities are rendered intelligible. These researchers view particulars of these situations as contributing to the creation of generic knowledge—not despite, but because of, these specificities.

Furthermore, these conditions of knowledge production also shape the contours of real-world implementations. The relationship between empirical studies and real-world applications is commonly understood through a representational model. That is, empirical

studies are often seen as scaled-down or simplified representations of the world, designed to generate knowledge that can be generalized. Researchers, being naturally aware of this representational nature, account for it by discussing the study's limitations (e.g., using the concept of ecological validity).

In contrast to this representational model, the STS scholarship highlights the performativity of laboratory settings (e.g., see discussions in [24, 33, 54]), showing that it is not only that the laboratories are deliberately constructed to be scaled models of the external world, but it is also this world itself that is often actively reconfigured to accommodate the products of laboratory-based technoscientific knowledge. Since the products of technoscience (such as robots) have been shown to be operational under laboratory conditions, there is also a reverse process of adjusting the real-world conditions to resemble the laboratory, as the two are never completely identical. Without such adjustments, technoscience may not function as it is studied in the labs. An example of this is how many contemporary self-driving cars require specific road infrastructure to function correctly—since testing of such vehicles is often conducted with assumptions of particular roads in mind—which, in turn, leads to the remaking of infrastructure to accommodate that [27, 45, 59]. Other examples of this reverse influence can be found in studies on smart homes and devices, which highlight how our homes are increasingly being remade to accommodate robots. For instance, Forlizzi [25] shows that for the robotic vacuum cleaner Roomba to function properly, users must tidy their homes beforehand, for such robots cannot navigate around certain items on the floor.

From this perspective, research methods are not simply technical instruments for collecting data or measuring outcomes. They are world-making practices, i.e., frameworks that participate in producing the very phenomena they claim to observe [90]. This performative view of method disrupts the traditional ideal of scientific detachment, drawing attention to the entanglement of epistemology, ethics, and politics in research. What constitutes legitimate data, what is discarded as an anomaly or noise, and what kinds of questions are posed in the first place are all shaped by the methodological assumptions and institutional settings in which research is conducted.

Crucially, methods do not remain confined to laboratories or controlled settings. They "travel" [55], carrying with them embedded assumptions about validity, generalizability, and objectivity into new contexts, often without sufficient reflection on their effects. The details of how and where methods operate are not obstacles to general knowledge, but they are the very mechanisms by which such knowledge is produced and legitimized. For our purposes, we refer to the methodological setups—such as experiments, surveys, workshops, or Wizard of Oz setups—that are frequently employed in the HRI scholarship [7, 8, 20, 47]. Reflecting critically on these conditions is important because this research also influences how robots will operate outside of research settings. While there will always be a difference between scientific research and real-world implementations, insofar as the design knowledge on robotic fallibility seeks to inform applications of robots in the wild, it is warranted to reflect on how conditions of knowledge production shape our repertoire of design responses to robot failures beyond the laboratory. Understanding methods as situated and generative enables us to ask not only how we study failure, but also how those studies shape what failure means in real-world contexts.

3.3 What is the purpose of knowledge?

The final question we pose invites reflection on the purposes and stakes of technoscientific inquiry. From a feminist technoscience perspective, knowledge is never a neutral, self-contained product. It is always situated within broader political, economic, and institutional agendas that guide what, how, and why something should be studied [78, 81, 82, 107, 120, 121]. Thus, scientific research is never conducted purely for the sake of satisfying researchers' curiosity but is part of a broader research agenda that is partially determined by industry, state, and other interests. These factors influence which research questions receive funding, which methods are valorized, and which applications are prioritized; as a result, specific topics, approaches, and themes gain visibility.

This does not negate scientific autonomy but highlights that external contexts always shape it. For example, Šabanović [87] demonstrated how Japanese roboticists actively incorporate local cultural norms as an epistemological justification for their work, relating their scientific and engineering research to specific Japanese values and practices. At the same time, this culturally aware positioning by Japanese roboticists, as Šabanović shows, also risks reproducing certain conservative values, precluding discussions on whether these values and practices are desirable for other social actors.

After specific questions are chosen, the scope of potential results is further limited by the way they are examined. For instance, the framing of a research problem, e.g., conditions of trust in robots or reducing failure rates, already presumes a specific imagined future in which robots play a particular societal role. This, in turn, may influence how research design is constructed and what design recommendations are provided, privileging certain ways of knowing robotic breakdown and not others. By shifting focus from knowledge as a purely epistemic achievement to knowledge as a world-making practice, feminist technoscience offers a way to hold ourselves accountable—not just for what we know, but for what that knowledge enables. It encourages design researchers to envision and pursue reflexive approaches to technological development. Scrutinizing the rationales for studying robot failures can thus help us, as design researchers, become reflective and aware of the assumptions that are either explicitly or implicitly part of our research agendas.

4 Robotic fallibility through a feminist lens: subjecthood, performativity of methods, and the research end goals

In the previous section, we introduced the research questions based on the feminist technoscience framework. In this section, we will examine the robotic fallibility scholarship through the lens of these questions. By 'robotic fallibility scholarship,' we understand papers published in HRI and HCI venues that directly address robot mistakes, failures, errors, and breakdowns as a source of design knowledge.

To construct our corpus, we began by identifying key literature reviews in HRI and HCI that survey failure- and error-related

research (e.g., [43, 62]). From these anchor points, we manually traced networks of citation and influence, following references both backward and forward to identify frequently cited and thematically relevant papers. This process revealed a body of work that treats robotic fallibility as more than a technical glitch, positioning it as a design opportunity, a user experience concern, or a site for critical reflection. We prioritized papers published in top-tier HRI and HCI venues (e.g., CHI, HRI, TOCHI, CSCW, DIS) that make robotic fallibility central to their empirical, theoretical, or methodological inquiry. We excluded works that only mention failure in passing or treat it purely as a technical performance metric. Instead, our focus is on studies where failure is central to the research question, interaction scenario, or conceptual framing.

The resulting analysis is best understood as a critical interpretive review, rather than a systematic or scoping review. Our goal is not comprehensive leverage, but rather a theory-driven engagement with how failure is framed, studied, and designed with. Rather than aiming for generalizability, our approach reveals the epistemological and ethical stakes embedded in a particular trajectory of HCI/HRI scholarship.

4.1 The subject of knowledge in robotic fallibility

We argue that the epistemic standpoint of designers–and by extension, design teams and companies–is systemically privileged in much robotic fallibility scholarship. Robotic breakdowns, errors, and failures are treated as phenomena to be interpreted and managed by designers in the service of improving interaction, performance, or user experience (e.g., [4, 58, 70, 84, 100, 127]. The designer is often positioned as the primary sense-maker, translating failures into actionable insights. While participatory approaches to studying failures exist, they remain scarce [109].

While the designer frequently appears as the privileged subject of knowledge, this position is rarely examined with critical nuance. This figure is often portrayed as a detached observer (in line with the notion of "seeing everything from nowhere" [37]) who studies how users encounter robotic failures, draw conclusions, and implement corrective or strategic measures. The designer is granted epistemic authority, while users are reduced to data sources, test subjects, or behavioral inputs. Even in cases where designers are portrayed as "user-centered," the asymmetry remains: it is the designer who is authorized to know, interpret, and act, while the user's role is passive and responsive. For instance, studies commonly use researcher-defined metrics, such as "error severity" scales, to categorize failures [30, 85, 115]. Moreover, users' responses to robot failures are often analyzed behavioristically, as a combination of data about their gaze, laughter, head movements, or other behavioral expressions [31, 73]. Yet the decision to see and interpret the user in these terms is itself an epistemic choice. As a result, the authority to define what constitutes a problem and what constitutes a solution is unevenly distributed.

On the other hand, in much of the robotic fallibility literature, the figure of the user is framed as a discrete individual—someone who interacts with robots, experiences breakdowns, and adapts accordingly. This framing frequently relies on demographic segmentation (e.g., age, gender, technical literacy) or user categories derived from experimental design (e.g., [52, 72, 89]). While useful for certain forms of HRI research, this view risks flattening the user into a predictable, stable actor, overlooking the diverse ways fallibility is encountered, interpreted, and responded to in real-world settings.

This limitation becomes especially visible in *collective settings*, such as care homes, schools, and hospitals. Here, robots operate within complex social ecologies where responses to failure are co-constructed by multiple actors. Ethnographic studies of elder care settings, for instance, show that when robots like PARO, Pepper, or Zora malfunction, e.g., failing to respond, initiate interactions at inappropriate moments, or behave unpredictably, it is often care workers, not residents, who reset robots, explain malfunctions, or soothe confused or distressed residents [16, 75]. These acts go beyond ad hoc fixes; they are forms of situated know-how and improvisational expertise, rooted in the routines, spatial arrangements, and affective textures of caregiving [61].

In these settings, robotic functionality depends on the invisible labor of care staff who accommodate technological fallibility. Fallibility becomes "located accountability" [105], diffused across a collective of actors who maintain the conditions for the robot to function meaningfully. Yet this collective labor often goes unrecognized in HRI discourses. A designer-centric epistemology persists, which privileges insights drawn from controlled experiments, system logs, and quantifiable feedback, while overlooking the rich, messy, and embodied forms of knowledge that emerge in situated contexts [86, 116].

Moreover, even when users are treated as individuals, they are often implicitly expected to perform a range of *affective and cognitive labor* in the face of robotic breakdowns. They are expected to remain patient, forgiving, and adaptive—to understand that robots are "still learning," to troubleshoot minor malfunctions, or to interpret ambiguous behavior generously. This is sometimes framed as "user resilience" [26] or "tolerance to failure" [52], but such framings may obscure the asymmetrical burden placed on users, especially those already socially positioned as caregivers or support figures. Care workers, parents, and educators are often expected to assume the emotional labor of maintaining the illusion of seamless interaction, compensating for technological limitations through their own attentiveness, patience, and emotional self-regulation. This emotional labor is rarely accounted for in system design or evaluation, yet it plays a crucial role in enabling robotic systems to function acceptably in practice [48]. What is at stake, then, is not only the usability of robotic systems, but also the reproduction of invisible and undervalued care labor in environments that are already ethically fraught. Recognizing this dimension of user experience requires expanding the scope of HRI and design research to include the social, emotional, and ethical frictions that arise when users are implicitly or explicitly enrolled as caretakers of robotic fallibility [22].

4.2 Methodological conditions and the making of robotic failure

Building on the feminist insight that methods are not neutral but world-making, we now turn to examine how this plays out in the

specific context of robotic fallibility scholarship. In HRI, robot failures are frequently studied through experimental setups designed to test user responses to malfunctions (e.g., [23, 58, 69, 73]). Centrally, we highlight that through such methods, the robotic fallibility is not only represented for scientific purposes but also performed in real-world contexts. The acknowledgement of this intimate connection between robotic fallibility *as* a product of scientific research and engineering practice and robotic fallibility *as* a set of material artifacts, systems, and practices, situated in the real world, can thus be a design concern.

This performative aspect raises new questions for the study of robotic fallibility. If research practices facilitate the materialization of real-world implementations, such practices may be significant to consider. When designers accept that robotic fallibility is not something to design around, but to design with, one might ask what relations with fallible robots emerge for everyone involved. For example, methodological considerations sometimes predispose researchers to remove certain users from the study sample. In a study examining the effects of failure severity on people's evaluations of robots, for instance, van Waveren et al. [115] had to exclude four participants because there was no guarantee that these individuals had noticed the experimental manipulation. This illustrates how methodological requirements—in this case, participants' awareness of robot failure as an instrumental experimental variable—become crucial for defining the boundaries of what constitutes failure and how it can be studied.

Other examples are concepts of error detection and mitigation that are widely used in the literature on robotic fallibility (e.g., [58, 99, 104, 111]). In one sense, they are scientific and epistemological devices that one employs to gain knowledge about the fallibility of robotics. Through this framework, one can create scientific models or representations conducive to, for example, building more robust explanations or deeper interpretations. At the same time, such a framework may have a role beyond its epistemological and scientific aspects. That is, it may correlate or correspond to certain practices in the world under certain conditions. To take error detection as an example, some scholars suggest leveraging social cues, such as various details about people's facial expressions, to design robots that can dynamically respond to their own failings by analyzing these social cues through machine learning techniques [12, 64, 108]. It should also be noted that there is a well-developed critique regarding the treatment of facial expressions as sources of reliable, computationally tractable information about inner emotional states [102, 103].

At the same time, accounting for the performative aspect of scientific knowledge requires that we further probe the strategy of using social cues for detection. Notwithstanding the respective merits or downsides of this approach in scientific terms, it is worth considering this error detection strategy in terms beyond its epistemological potential. It does not seem far-fetched to imagine how, if implemented, this design strategy may bring to the surface an array of ethical and social questions. If robots are designed to recognize people's social cues as indicators of the functionality of their behavior, can it potentially imply a certain form of problematic surveillance of people's behavior? This is not only a matter of potential privacy violations as infringement of individual autonomy, but also of the broader structuring role of technology within the relationships in which people socialize, work, and care for one another. Depending on the setting of implementation—whether it is a workplace, public space, or a healthcare setting—these questions may multiply.

This is not only a hypothetical concern. For instance, Yu et al. [126] document how delivery robots misinterpret sidewalk interactions due to simplified assumptions about human spatial behavior, leading to user confusion and infrastructural friction. These breakdowns stem directly from prior design assumptions based on constrained experimental models of "failure" and "recovery". Similarly, Kureha [53] describes multiple examples where the robot apologies, often studied in controlled lab settings, fail to produce the intended effect in dynamic public environments, where the social context differs significantly.

These examples illustrate how methodological framings, such as predefined failure types or fixed interaction scripts, do not merely measure fallibility but shape how it manifests and is acted upon in deployment. They follow a lineage of historical evidence of a performative connection between research on technology and settings where technology is implemented [25, 54, 55, 59]. If studies of robotic fallibility inform real-world deployments, and if those deployments, in turn, reshape the spaces in which robots operate, then the methods of knowing fallibility are themselves design choices with downstream consequences. For this reason, we encourage HRI and HCI researchers to attend not only to the epistemic strengths of their methods, but also to their sociopolitical effects. The move to design with fallibility must also include critical reflection on how the conditions under which the fallible robots are studied shape the possible futures in which fallible robots are expected to function.

4.3 The ends of robotic fallibility research

Studying robotic fallibility does not occur in a vacuum but is always motivated by specific research agendas stemming from robotics' relationship with social actors, such as governments, industry, and research institutions. In this section, we explore the epistemic consequences of these agendas, highlighting how they shape which fallibilities are seen as worth addressing, and how.

While focusing only on aspects or occurrences of robotic fallibility, as opposed to trying to address all of them at once, is an unavoidable part of the design research scoping, we consider it valuable to explicate these rationales and reflect on them. Across the literature, we identify two primary motivations underlying this research. First, robotic fallibility is perceived as a problem because it *compromises user experience, thereby decreasing the acceptance of technology*. This can manifest when interaction interruptions, resulting from the robotic errors, cause frustration for a user, leading them to use the robot less in the future (e.g. [43, 58, 73]). This rationale is also connected to the discussions of trust and (over)reliance in human-robot relations and the impact that robotic errors may have on it (e.g. [30, 111, 127]).

Second, robotic fallibility is seen as an issue to be addressed because it can potentially *impede the possible economic benefits* that a certain actor employing robots may yield. For example, an increased rate of robotic errors may contribute to customer retention [72] or lead to an increase in customer support costs [43]. Such a

close coupling of robotic fallibility and economic rationales may inadvertently obscure any robotic fallibilities, the mitigation of which does not yield tangible economic benefits. Second, even in scenarios where such robotic fallibilities are not obscured but acknowledged, framing the research agenda in this way may motivate social actors to address these fallibilities precisely through the lens of economic value. These two mechanisms shape the possible ranges of design strategies for fallible robotics, at the risk of always prioritizing cost-efficient ones over others. As a result, other concerns, such as sustainability or perspectives of non-users (non-customers or bystanders), may fall out of scope.

While these two motivations do not define the entire field, they structure much of the design agenda. If left unexamined, they risk narrowing how robotic fallibility is understood and addressed. It is in this context that we raise a broader concern: when fallibility is addressed for the purposes of improving user experience or having economic gains, it often reinforces a uniform ideal of what a "successful" (as opposed to "failed") robot is; implicitly framing failure as deviation from this ideal, without recognizing the underlying normative commitments. Critical cultural theory demonstrates that interpretations of who fails, what constitutes failure, and how failure occurs are inseparable from normative assumptions about how society ought to be organized [3, 14, 35]. For instance, borrowing from theories in crip technoscience, researchers have demonstrated how ableist assumptions are reproduced in the design of robots, resulting in users sometimes experiencing their bodies as 'failing' to adhere to an underlying, technically embedded vision of normality, rather than the other way around [34, 36, 125]. Treusch [113] has leveraged queer theories of failure to recenter the affective labor that occurs in the robotics lab, as exemplified by researchers laughing when the constraints of robots are exposed. Similarly, Harrison et al. [40] call for expanding studies of robotic fallibility to include not just technical or quantitative analyses, but also the affective and interpretive dimensions involved in encountering breakdown. By interweaving theoretical, methodological, and ethical questions, they explore how we can "reconcile the notion that failure is a socioculturally specific judgment with the roboticists' view of robotic failure as technical breakdown" [40].

These perspectives underscore the inherent value-laden nature of the design process. Addressing, fixing, neglecting, or even leveraging a particular robotic failure in a certain manner implies a commitment to one vision of success over another. Given that no single definition of robotic success is universally valid, these normative underpinnings deserve more explicit attention. This becomes especially clear when looking at real-world deployments. For example, Starship recently deployed delivery robots on various U.S. university campuses. Designed to operate seamlessly in pedestrian environments, these robots often prompted unexpected public responses: students started a petition against the robots [128], used them in pranks [129], or anthropomorphized their behavior [130]. These interactions were not simply irrational or disruptive; they were situated, interpretive acts through which people actively negotiated the robot's presence in shared space. What is at stake here is not only public acceptance but also questions of who feels entitled to shape shared spaces, and whose discomfort or resistance is dismissed as unproductive or anti-innovation.

Although some HRI studies do examine socially salient aspects of robot failure (e.g., [110, 111]), they often share an assumption that there is a single, internally coherent, and temporarily stable definition of 'failure' to be scientifically discovered in 'the social' or 'interaction.' Not only can there be many contrasting, sometimes conflicting definitions held by different stakeholders, but these interpretations rarely carry the same epistemic weight within formal design processes as those of designers or researchers. Exploring differences reveals that robotic failure and success are not self-evident categories, but rather ones that are shaped by competing values, interests, and assumptions. These tensions emphasize design being a "site of contestation" [21]—a space where users, publics, and other stakeholders can reinterpret robotic behavior and challenge the assumptions embedded in design decisions. Acknowledging this opens space for more reflexive, inclusive, and socially responsive approaches to the design of fallible robots.

Moreover, we consider it critical to scrutinize the epistemological standpoint of seeing all robot failures as productive opportunities for improvement. This risks framing any robot as univocally good and infinitely improvable artifact, and risks positioning social resistance, discomfort, or hesitation as always mere misunderstandings of the robot's design and purpose (on critique and resistance to robotics see [2, 32, 71, 81, 107, 124]).

On the contrary, under certain circumstances, failures may indicate a robot's unsuitability for particular contexts, tasks, or users, rather than a temporal mishap that can be overcome with time. This criticism does not suggest a return to framing all failures in HRI as a problem to be avoided but calls for finer distinctions between different types of failure: those that can be generatively leveraged, and those that may warrant reconsideration or even rejection of robotic deployment. So far, only failures posing immediate threats to physical safety tend to be viewed as categorically unacceptable.

5 Discussion: design considerations and ways forward for studying robotic fallibility

In this paper, we have explored how robotic fallibility is framed, studied, and addressed within HRI, highlighting the ethical, social, and political implications of these approaches. Treating failures, errors, and breakdowns of robots not merely as obstacles to be eliminated but as meaningful design materials is itself a significant and value-laden design decision. Designing with, rather than against, fallibility entails a myriad of interpretive choices, such as what counts as a failure, how it should be addressed, and what outcomes are desirable. These choices, as we have shown, are never purely technical but are shaped by normative assumptions, disciplinary norms, and sociopolitical contexts. Therefore, they are contingent rather than inevitable decisions about where to draw the line between 'success' and 'failure.'

This work, too, is situated. Our analysis emerges from within particular disciplinary and institutional contexts shaped by the epistemic cultures of HRI, design research, and feminist technoscience. The questions we ask, the concepts we prioritize, and the critiques we offer are made possible by the research infrastructures we are embedded in, including the affordances and constraints of academic publishing, research funding, and collaborative networks. We acknowledge that our perspective is shaped by specific

methodological commitments and theoretical orientations, positioning us within broader conversations about design, technology, and ethics. Rather than treating this as a limitation, we understand it as a necessary part of accountable knowledge-making. Recognizing the situatedness of our knowledge, we offer this work as one contribution among many—shaped by context, responsive to its conditions, and open to reinterpretation by others working from different perspectives.

In this final section, we offer considerations for more reflexive and inclusive approaches to designing with robotic fallibility. Based on our previous analysis, we synthesize three key recommendations: (1) considering the social ambiguity of failure and the epistemic agency of different stakeholders; (2) reflecting on the performativity of research methods and their broader consequences; (3) interrogating the underlying motivations that shape research agendas in HRI and design.

5.1 The social ambiguity of failure and the epistemic agency of different stakeholders

We urge designers to further reflect on agency in designing with robotic fallibility. In Section 3.1, we discussed how the construction of agency within technoscientific projects imprints on the characteristics of the knowledge claims produced. Section 4.1. has explored how the designer often remains the privileged subject in deciding what counts as robot failure. While not denouncing the scientific value of technical and engineering definitions of failure, error, and breakdown, we nonetheless urge the design research community to consider the social ambiguity of robotic fallibility.

For example, as we discussed in Section 4.3, the diverse uses of Starship delivery robots do not qualify as 'failures' in a technical sense. Instead, they reveal the *social ambiguity of robotic fallibility*—how the same event might be interpreted as a breakdown, a joke, a publicity stunt, or a moment of play. By that, we refer to how robotic fallibility may change in time and be subject to various interpretations, depending on the context. Robotic fallibility, in other words, is also a matter of human experience, rather than only an observable and measurable event or property in the world, as it is often defined. The designers' responses to the users' negotiations with Starship robots primarily involved adding lights, polite voice prompts, or signals to indicate the robot's intent [119]. These addressed surface-level interaction issues but left deeper normative and social dynamics largely unexamined. As a result, the success of these robots becomes narrowly aligned with operational autonomy, logistical efficiency, and economic gains, while broader sociotechnical frictions remain outside the scope of design concern. Thinking of users primarily in terms of observable behaviors also risks precluding other forms of design engagement.

Moreover, encountering robotic fallibility also comes with emotional labor that often remains underrecognized in the design research—further studies could consider these emotions not just as a scientific resource, but also as a relational, collective, and lived experience of participants, both in the lab and beyond. This recommendation aligns with feminist and care-centered perspectives in HCI that emphasize maintenance, relationality, and the uneven distribution of labor in technological systems (e.g. [17, 48, 49]). Understanding robotic fallibility as a shared, interpretive, and emotional experience necessitates expanding the design lens to encompass practices of care, adaptation, and resistance. One direction to pursue in this respect, following Lee et al. [57], could be finding ways to recenter the workers' perspectives in research, so that their labor in addressing robot failures does not get undervalued.

Bringing this labor into view allows us to rethink what constitutes valuable user knowledge. It shows that users are not passive recipients of technology, but active participants in making robots workable. Moreover, it challenges the assumption that fallibility is something that only designers understand and resolve. Instead, it becomes clear that collective, relational knowledge practices—practices of care, coordination, and adaptation—are central to how robotic fallibility is navigated and made meaningful. This recommendation extends participatory and reflective design traditions in HCI, which argue that the authority to define problems and shape systems should be shared with those most affected by them (e.g.) [11, 80, 92]. We argue that robotic failure, too, must be understood as a contested category—co-produced through situated interpretations, emotional labor, and sociotechnical relations.

5.2 The performativity of research methods

In Sections 3.2 and 4.2, we proposed that the methods for studying robotic fallibility not only represent it but also actively shape how it will exist in the world outside of design labs and research settings. Rather than being exclusively scientific instruments of knowledge-making, the role of such methods also extends to shaping the fallible robots that users will interact with. Some design implications result from these theoretical considerations.

We suggest that research methods can be viewed as design interventions insofar as they provide scaffolding for how fallible robots will operate in the wild. Insofar as the performativity argument holds, the concepts such as detection or recovery are not only scientific concepts but also potentially concrete social and material techniques that would have to materialize in some form if the research direction on fallible robots gains traction. Examining the various social, ethical, and political implications of these methods from this perspective may then be a form of much-needed designerly inquiry.

Realizing this form of reflexivity in practice may take different forms. We suggest two strategies, *documentation* and *reverse translation*, as points of inspiration. The first one refers to the efforts of the research and design teams to discuss and document the methodological assumptions of their studies in ways that go beyond seeing them as only epistemological limitations, such as concerns for ecological validity. Instead, these forms of documentation may entail cataloging and reporting the effects that these assumptions may have on real-world implementations. One potential type of object for such analysis is the emerging datasets on robot errors [13, 51]. Our second suggestion, reverse translation, proposes the opposite move. Design researchers can start by imagining the desirable kind of robotic fallibility and then consider the research frameworks, concepts, and methods that may correspond to the project of producing knowledge about that kind of desirable fallibility. As HRI research continues to expand beyond its traditionally quantitative and empirical foundations—opening itself to qualitative [116], designerly

[68], and critical approaches [66]—this kind of epistemological and methodological reasoning becomes particularly timely. Such an exercise acknowledges the normativity of designing with robotic fallibility and may help both surface it and leverage these insights for more expansive and generative design research.

5.3 The underlying motivations behind the design research

In Sections 3.3 and 4.3, we argued that the design research on robotic fallibility may benefit from recognizing how its research agendas are always already intertwined with external contexts and actors, shaping the purposes and end goals of studies. Similarly, in its goal not to take robotics for granted, critical HRI examines the goals and directions of such research, questioning the fundamental purpose of robots in society [65, 93]. Aligning with this scholarship, we suggest that the distinction between a 'successful' and 'failed' robot is also a normative, potentially conflict-laden demarcation, rather than only a descriptive, neutral gesture.

Moreover, if every failure is taken as a call to improve the robot, then user withdrawal, skepticism, or critique may be too easily dismissed as resistance to be overcome. Always increasing user trust in robots is normatively desirable only if the assumption that robots are universally good holds. As this may not be necessarily the case, it is worthwhile for the HRI/HCI scholars to consider in which situations the failure of the robot may be an indication of well-placed user mistrust in the robot, making a case for not designing the robot at all and respecting that resistance to robots can be legitimate.

For this, future studies of robotic fallibilities can draw on HCI scholarship on "unmaking" [74, 98] and "undesigning" [79], which questions the primacy of creating novel artifacts, both in theory and practice [74, 98]. Against the imperative of always creating new things, unmaking pays attention to the afterlives of design practice, helping us work through both materials and human-technology relations in cases of robot failures that do not necessitate repair [60]. Rather than positioning design as inherently constructive, unmaking and undesigning allows for the *intentional negation of design*—asking when it is better not to build, not to fix, or to let technologies decline. This perspective is particularly useful when reflecting on robotic failure, where repair is not always the most appropriate response, and where disengagement may be both ethically and socially productive.

Lastly, keeping in mind the risks of focusing exclusively on economic rationales can help to remain sensitized to other concerns, such as sustainability or (non-)users' perspectives. Failures are often framed as friction points between system and user, but they also have environmental, labor, and infrastructural consequences, as we hope to have demonstrated in this paper. Designing with fallibility includes considering the long-term and societal impacts of robotic systems, such as repair cycles and the labor required to maintain seamless interactions, especially when these costs are unevenly distributed. We urge design researchers to remain critical of their underlying research motivations and assumptions when designing studies, writing grant applications, assembling research teams, talking with stakeholders, and organizing events. Cultivating this sensibility does not mean rejecting economic considerations altogether but resisting their dominance and making room for ethically grounded, context-sensitive, and inclusive approaches to robot design.

6 Conclusion

There will be failures, breakdowns, and errors in human-robot interaction; designing with them can be a means to acknowledge these realities. At the same time, this shift merits not only a new scientific and design research agenda, but also epistemological, ethical, and political discussion of its implications. In this paper, drawing on knowledge from the feminist technoscience scholarship, we explored how questioning agency, the performative role of methods, and the motivations behind research can help to reflect on the underlying assumptions of this research field. Insofar as design is also a world-making practice, thinking about who, how, and why designers engage in research is a valuable contribution. We encourage HRI and HCI researchers to consider the fallibility of robots and other design artifacts, not only as a learning opportunity and a material for design inquiry, but also as a means to understand how failure is relationally experienced, methodologically enacted, and further repurposed for particular goals.

Acknowledgments

This research was carried out within the AI DeMoS Lab funded by the TU Delft AI Initiative.

References

[1] Alison Adam. 1998. *Artificial Knowing: Gender and the Thinking Machine*. Routledge, London.

[2] Tanja Ahlin and Anna Mann. 2025. Ambiguous animals, ambivalent carers and arbitrary care collectives: Re-theorizing resistance to social robots in healthcare. *Social Science & Medicine* 365: 117587. https://doi.org/10.1016/j.socscimed.2024.117587

[3] Arjun Appadurai and Neta Alexander. 2020. *Failure*. Polity press, Cambridge Medford, Mass.

[4] Markus Bajones, Astrid Weiss, and Markus Vincze. 2016. Help, Anyone? A User Study For Modeling Robotic Behavior To Mitigate Malfunctions With The Help Of The User. https://doi.org/10.48550/ARXIV.1606.02547

[5] Shaowen Bardzell. 2010. Feminist HCI: taking stock and outlining an agenda for design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1301–1310. https://doi.org/10.1145/1753326.1753521

[6] Shaowen Bardzell and Jeffrey Bardzell. 2011. Towards a feminist HCI methodology: social science, feminism, and HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 675–684. https://doi.org/10.1145/1978942.1979041

[7] Christoph Bartneck. 2017. Reviewers' scores do not predict impact: bibliometric analysis of the proceedings of the human–robot interaction conference. *Scientometrics* 110, 1: 179–194. https://doi.org/10.1007/s11192-016-2176-y

[8] Paul Baxter, James Kennedy, Emmanuel Senft, Severin Lemaignan, and Tony Belpaeme. 2016. From characterising three years of HRI to methodology and reporting recommendations. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 391–398. https://doi.org/10.1109/HRI.2016.7451777

[9] Arne Berger, Albrecht Kurze, Andreas Bischof, Jesse Josua Benjamin, Richmond Y. Wong, and Nick Merrill. 2023. Accidentally Evil: On Questionable Values in Smart Home Co-Design. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–14. https://doi.org/10.1145/3544548.3581504

[10] Jessica Bley, Alexander Eriksson, Lisa Johansson, and Mikael Wiberg. 2023. Design friction in autonomous drive—exploring transitions between autonomous and manual drive in non-urgent situations. *Personal and Ubiquitous Computing* 27, 6: 2291–2305. https://doi.org/10.1007/s00779-023-01780-7

[11] Susanne Bødker and Morten Kyng. 2018. Participatory Design that Matters—Facing the Big Issues. *ACM Transactions on Computer-Human Interaction* 25, 1: 1–31. https://doi.org/10.1145/3152421

[12] Alexandra Bremers, Alexandria Pabst, Maria Teresa Parreira, and Wendy Ju. 2024. Using Social Cues to Recognize Task Failures for HRI: Overview, State-of-the-Art, and Future Directions. https://doi.org/10.48550/arXiv.2301.11972

[13] Alexandra Bremers, Maria Teresa Parreira, Xuanyu Fang, Natalie Friedman, Adolfo Ramirez-Aristizabal, Alexandria Pabst, Mirjana Spasojevic, Michael Kuniavsky, and Wendy Ju. 2023. The Bystander Affect Detection (BAD) Dataset for Failure Detection in HRI. In *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 11443–11450. https://doi.org/10.1109/IROS55552.2023.10342442

[14] Lauren E Bridges. 2021. Digital failure: Unbecoming the "good" data subject through entropic, fugitive, and queer data. *Big Data & Society* 8, 1: 205395172097788. https://doi.org/10.1177/2053951720977882

[15] David Cameron, Stevienna De Saille, Emily C. Collins, Jonathan M. Aitken, Hugo Cheung, Adriel Chua, Ee Jing Loh, and James Law. 2021. The effect of social-cognitive recovery strategies on likability, capability and trust in social robots. *Computers in Human Behavior* 114: 106561. https://doi.org/10.1016/j.chb.2020.106561

[16] Felix Carros, Isabel Schwaninger, Adrian Preussner, Dave Randall, Rainer Wieching, Geraldine Fitzpatrick, and Volker Wulf. 2022. Care Workers Making Use of Robots: Results of a Three-Month Study on Human-Robot Interaction within a Care Home. In *CHI Conference on Human Factors in Computing Systems*, 1–15. https://doi.org/10.1145/3491102.3517435

[17] David Chatting. 2023. Automated Indifference. *Interactions* 30, 2: 22–26. https://doi.org/10.1145/3580299

[18] Nazli Cila, Maria Luce Lupetti, Luciano Cavalcante Siebert, and Janna Van Grunsven. 2025. Dramatic Things: Investigating Value Conflicts in Smart Home through Enactment and Co-speculation. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 1–17. https://doi.org/10.1145/3706598.3713138

[19] Lorraine Code. 1991. *What can she know? feminist theory and the construction of knowledge*. Cornell University Press, Ithaca.

[20] Kerstin Dautenhahn. 2018. Some Brief Thoughts on the Past and Future of Human-Robot Interaction. *ACM Transactions on Human-Robot Interaction* 7, 1: 1–3. https://doi.org/10.1145/3209769

[21] Carl DiSalvo. 2022. *Design as democratic inquiry: putting experimental civics into practice*. The MIT Press, Cambridge, Massachusetts.

[22] Anna Dobrosovestnova, Isabel Schwaninger, and Astrid Weiss. 2022. With a Little Help of Humans. An Exploratory Study of Delivery Robots Stuck in Snow. In *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 1023–1029. https://doi.org/10.1109/RO-MAN53752.2022.9900588

[23] Kanghui Du, Dražen Brščić, and Takayuki Kanda. 2025. Don't Just Stop Here! Human-Inspired Solutions for Sudden Robot Stops. In *2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 293–302. https://doi.org/10.1109/HRI61500.2025.10973836

[24] James Evans and Andrew Karvonen. 2014. 'Give Me a Laboratory and I Will Lower Your Carbon Footprint!' - Urban Laboratories and the Governance of Low-Carbon Futures: Governance of low carbon futures in Manchester. *International Journal of Urban and Regional Research* 38, 2: 413–430. https://doi.org/10.1111/1468-2427.12077

[25] Jodi Forlizzi. 2007. How robotic products become social products: an ethnographic study of cleaning in the home. In *Proceedings of the ACM/IEEE international conference on Human-robot interaction*, 129–136. https://doi.org/10.1145/1228716.1228734

[26] Knut Fossum and Abdul Basit Mohammad. 2015. Approaching Human-Robot Interaction with Resilience. In *Space Safety is No Accident*, Tommaso Sgobba and Isabelle Rongier (eds.). Springer International Publishing, Cham, 295–302. https://doi.org/10.1007/978-3-319-15982-9_35

[27] Matthew Franchi, Maria Teresa Parreira, Fanjun Bu, and Wendy Ju. 2025. The Robotability Score: Enabling Harmonious Robot Navigation on Urban Streets. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 1–17. https://doi.org/10.1145/3706598.3714009

[28] William Gaver, John Bowers, Tobie Kerridge, Andy Boucher, and Nadine Jarvis. 2009. Anatomy of a failure: how we knew when our design went wrong, and what we learned from it. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2213–2222. https://doi.org/10.1145/1518701.1519040

[29] Petra Gemeinboeck. 2021. The Aesthetics of Encounter: A Relational-Performative Design Approach to Human-Robot Interaction. *Frontiers in Robotics and AI* 7: 577900. https://doi.org/10.3389/frobt.2020.577900

[30] Romi Gideoni, Shanee Honig, and Tal Oron-Gilad. 2024. Is It Personal? The Impact of Personally Relevant Robotic Failures (PeRFs) on Humans' Trust, Likeability, and Willingness to Use the Robot. *International Journal of Social Robotics* 16, 6: 1049–1067. https://doi.org/10.1007/s12369-022-00912-y

[31] Manuel Giuliani, Nicole Mirnig, Gerald Stollnberger, Susanne Stadler, Roland Buchner, and Manfred Tscheligi. 2015. Systematic analysis of video data from different human–robot interaction studies: a categorization of social signals during error situations. *Frontiers in Psychology* 6. https://doi.org/10.3389/fpsyg.2015.00931

[32] Murray Goulden and Lewis Cameron. 2025. The role of mundane resistance in the spectacular failure of the smart home. *Big Data & Society* 12, 3: 20539517251361120. https://doi.org/10.1177/20539517251361120

[33] Matthias Gross. 2016. Give Me an Experiment and I Will Raise a Laboratory. *Science, Technology, & Human Values* 41, 4: 613–634. https://doi.org/10.1177/0162243915617005

[34] Josh Guberman and Oliver Haimson. 2023. Not robots; Cyborgs — Furthering anti-ableist research in human-computer interaction. *First Monday*. https://doi.org/10.5210/fm.v28i1.12910

[35] Jack Halberstam. 2011. *The Queer Art of Failure*. Duke University Press.

[36] Aimi Hamraie and Kelly Fritsch. 2019. Crip Technoscience Manifesto. *Catalyst: Feminism, Theory, Technoscience* 5, 1: 1–33. https://doi.org/10.28968/cftt.v5i1.29607

[37] Donna Haraway. 1988. Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies* 14, 3: 575. https://doi.org/10.2307/3178066

[38] Sandra Harding. 2015. *Objectivity and Diversity: Another Logic of Scientific Research*. University of Chicago Press, Chicago. https://doi.org/10.7208/chicago/9780226241531.001.0001

[39] Katherine Harrison, Giulia Perugia, Filipa Correia, Kavyaa Somasundaram, Sanne Van Waveren, Ana Paiva, and Amy Loutfi. 2023. The Imperfectly Relatable Robot: An Interdisciplinary Workshop on the Role of Failure in HRI. In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, 917–919. https://doi.org/10.1145/3568294.3579952

[40] Katherine Harrison, Kavyaa Somasundaram, and Amy Loutfi. 2025. The Imperfectly Relatable Robot: An Interdisciplinary Approach to Failures in Human–Robot Relations. In *How the robot made me feel*, Ericka Johnson (ed.). The MIT Press, Cambridge, 141–164.

[41] Adrian Hazzard, Chris Greenhalgh, Maria Kallionpaa, Steve Benford, Anne Veinberg, Zubin Kanga, and Andrew McPherson. 2019. Failing with Style: Designing for Aesthetic Failure in Interactive Performance. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–14. https://doi.org/10.1145/3290605.3300260

[42] Bart Hengeveld and Jordy Rooijakkers. 2019. Adding Friction to Frictionless Payments: A Haptic Interface for Digital Transactions. In *Proceedings of the Thirteenth International Conference on Tangible, Embedded, and Embodied Interaction*, 243–250. https://doi.org/10.1145/3294109.3300999

[43] Shanee Honig and Tal Oron-Gilad. 2018. Understanding and Resolving Failures in Human-Robot Interaction: Literature Review and Model Development. *Frontiers in Psychology* 9: 861. https://doi.org/10.3389/fpsyg.2018.00861

[44] Noura Howell, Audrey Desjardins, and Sarah Fox. 2021. Cracks in the Success Narrative: Rethinking Failure in Design Research through a Retrospective Trioethnography. *ACM Transactions on Computer-Human Interaction* 28, 6: 1–31. https://doi.org/10.1145/3462447

[45] Maya Indira Ganesh. 2024. *Auto-Correct: The Fantasies and Failures of AI, Ethics, and the Driverless Car*. ArtEZ Press.

[46] Steven J. Jackson and Laewoo Kang. 2014. Breakdown, obsolescence and reuse: HCI and the art of repair. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 449–458. https://doi.org/10.1145/2556288.2557332

[47] Malte Jung and Pamela Hinds. 2018. Robots in the Wild: A Time for More Robust Theories of Human-Robot Interaction. *ACM Transactions on Human-Robot Interaction* 7, 1: 1–5. https://doi.org/10.1145/3208975

[48] Waki Kamino, Selma Šabanović, and Malte F Jung. 2025. "A Robot's Life is Over When People Give Up": Socio-Technical Infrastructure for Sustaining Consumer Robots. In *2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 142–151. https://doi.org/10.1109/HRI61500.2025.10974029

[49] Cayla Key, Fiona Browne, Nick Taylor, and Jon Rogers. 2021. Proceed with Care: Reimagining Home IoT Through a Care Perspective. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–15. https://doi.org/10.1145/3411764.3445602

[50] Os Keyes and Kathleen A Creel. 2022. Artificial Knowing Otherwise. *Feminist Philosophy Quarterly* 8.

[51] Parag Khanna, Andreas Naoum, Elmira Yadollahi, Mårten Björkman, and Christian Smith. 2025. REFLEX Dataset: A Multimodal Dataset of Human Reactions to Robot Failures and Explanations. In *2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 1032–1036. https://doi.org/10.1109/HRI61500.2025.10974185

[52] Kim Klüber and Linda Onnasch. 2022. When Robots Fail—A VR Investigation on Caregivers' Tolerance towards Communication and Processing Failures. *Robotics* 11, 5: 106. https://doi.org/10.3390/robotics11050106

[53] Makoto Kureha. 2024. On the moral permissibility of robot apologies. *AI & SOCIETY* 39, 6: 2829–2839. https://doi.org/10.1007/s00146-023-01782-2

[54] Bruno Latour. 1983. Give Me a Laboratory and I will Raise the World. In *Science observed: perspectives on the social study of science*, Karin Knorr-Cetina and Mike Mulkay (eds.). Sage Publications, London, 141–170.

[55] John Law. 2004. *After method: mess in social science research*. Routledge, London New York. https://doi.org/10.4324/9780203481141

[56] Steven Lawrence, Melanie Jouaiti, Jesse Hoey, Chrystopher L. Nehaniv, and Kerstin Dautenhahn. 2025. The Role of Social Norms in Human–Robot Interaction: A Systematic Review. *ACM Transactions on Human-Robot Interaction* 14, 3: 1–44. https://doi.org/10.1145/3722120

[57] Hee Rin Lee, Sarah Fox, EunJeong Cheon, and Samantha Shorey. 2025. Minding the Stop-gap: Attending to the "Temporary," Unplanned, and Added Labor of Human-Robot Collaboration in Context. In *Proceedings of the 2025 ACM/IEEE International Conference on Human-Robot Interaction* (HRI '25), 34–44.

[58] Min Kyung Lee, Sara Kiesler, Jodi Forlizzi, Siddhartha Srinivasa, and Paul Rybski. 2010. Gracefully mitigating breakdowns in robotic services. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 203–210. https://doi.org/10.1109/HRI.2010.5453195

[59] Paul M. Leonardi. 2010. From Road to Lab to Math: The Co-evolution of Technological, Regulatory, and Organizational Innovations for Automotive Crash Testing. *Social Studies of Science* 40, 2: 243–274. https://doi.org/10.1177/0306312709346079

[60] Kristina Lindström and Åsa Ståhl. 2020. Un/Making in the Aftermath of Design. In *Proceedings of the 16th Participatory Design Conference 2020 - Participation(s) Otherwise - Volume 1*, 12–21. https://doi.org/10.1145/3385010.3385012

[61] Benjamin Lipp. 2023. Caring for robots: How care comes to matter in human-machine interfacing. *Social Studies of Science* 53, 5: 660–685. https://doi.org/10.1177/03063127221081446

[62] Benjamin Lipp. 2024. Robot Drama: Investigating Frictions between Vision and Demonstration in Care Robotics. *Science, Technology, & Human Values* 49, 2: 318–343. https://doi.org/10.1177/01622439221120118

[63] Dewen Liu, Changfei Li, Jieqiong Zhang, and Weidong Huang. 2023. Robot service failure and recovery: Literature review and future directions. *International Journal of Advanced Robotic Systems* 20, 4: 17298806231191606. https://doi.org/10.1177/17298806231191606

[64] Shannon Liu, Maria Teresa Parreira, and Wendy Ju. 2025. "I'm Done": Describing Human Reactions to Successive Robot Failure. In *2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 1458–1462. https://doi.org/10.1109/HRI61500.2025.10974098

[65] Sara Ljungblad and Mafalda Gamboa. 2024. Critical Perspectives in Human–Robot Interaction Design. In *Designing Interactions with Robots*. Chapman and Hall/CRC.

[66] Sara Ljungblad, Sofia Serholt, Tijana Milosevic, Niamh Ni Bhroin, Rikke Toft Nørgård, Pamela Lindgren, Charles Ess, Wolmet Barendregt, and Mohammad Obaid. 2018. Critical robotics: exploring a new paradigm. In *Proceedings of the 10th Nordic Conference on Human-Computer Interaction*, 972–975. https://doi.org/10.1145/3240167.3240267

[67] Helen E. Longino. 1990. *Science as social knowledge: values and objectivity in scientific inquiry*. Princeton University Press, Princeton, N.J.

[68] Maria Luce Lupetti, Cristina Zaga, and Nazli Cila. 2021. Designerly Ways of Knowing in HRI: Broadening the Scope of Design-oriented HRI Through the Concept of Intermediate-level Knowledge. In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, 389–398. https://doi.org/10.1145/3434073.3444668

[69] Akihiro Maehigashi, Kenta Kubo, Yun Nungduk, and Seiji Yamada. 2025. Effects of Robot Bowing during Apology on Trust Repair. In *2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 1478–1482. https://doi.org/10.1109/HRI61500.2025.10973849

[70] Amama Mahmood, Jeanie W Fung, Isabel Won, and Chien-Ming Huang. 2022. Owning Mistakes Sincerely: Strategies for Mitigating AI Errors. In *CHI Conference on Human Factors in Computing Systems*, 1–11. https://doi.org/10.1145/3491102.3517565

[71] Arne Maibaum, Andreas Bischof, Jannis Hergesell, and Benjamin Lipp. 2022. A critique of robotics in health care. *AI & SOCIETY* 37, 2: 467–477. https://doi.org/10.1007/s00146-021-01206-z

[72] Nika Meyer, Melanie Schwede, Maik Hammerschmidt, and Welf Hermann Weiger. 2022. Users taking the blame? How service failure, recovery, and robot design affect user attributions and retention. *Electronic Markets* 32, 4: 2491–2505. https://doi.org/10.1007/s12525-022-00613-4

[73] Nicole Mirnig, Gerald Stollnberger, Markus Miksch, Susanne Stadler, Manuel Giuliani, and Manfred Tscheligi. 2017. To Err Is Robot: How Humans Assess and Act toward an Erroneous Social Robot. *Frontiers in Robotics and AI* 4: 21. https://doi.org/10.3389/frobt.2017.00021

[74] Iohanna Nicenboim, Marie Louise Juul Søndergaard, Joseph Lindley, Anuradha Reddy, Yolande Strengers, Johan Redström, and Elisa Giaccardi. 2024. Unmaking-with AI: Tactics for Decentering through Design. *ACM Transactions on Computer-Human Interaction* 31, 6: 1–20. https://doi.org/10.1145/3685275

[75] Richard Paluch and Claudia Müller. 2022. "That's Something for Children": An Ethnographic Study of Attitudes and Practices of Care Attendants and Nursing Home Residents Towards Robotic Pets. *Proceedings of the ACM on Human-Computer Interaction* 6, GROUP: 1–35. https://doi.org/10.1145/3492850

[76] Soomi Park, Patrick G. T. Healey, and Antonios Kaniadakis. 2021. Should Robots Blush? In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–14. https://doi.org/10.1145/3411764.3445561

[77] Hannah R.M. Pelikan, Fanjun Bu, and Wendy Ju. 2025. The People Behind the Robots: How Wizards Wrangle Robots in Public Deployments. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 1–21. https://doi.org/10.1145/3706598.3713237

[78] Thao Phan, Jake Goldenfein, Monique Mann, and Declan Kuch. 2022. Economies of Virtue: The Circulation of 'Ethics' in Big Tech. *Science as Culture* 31, 1: 121–135. https://doi.org/10.1080/09505431.2021.1990875

[79] James Pierce. 2012. Undesigning technology: considering the negation of design by design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 957–966. https://doi.org/10.1145/2207676.2208540

[80] Xiang Qi and Junnan Yu. 2025. Participatory Design in Human-Computer Interaction: Cases, Characteristics, and Lessons. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 1–26. https://doi.org/10.1145/3706598.3713436

[81] Kathleen Richardson. 2016. Sex Robot Matters: Slavery, the Prostituted, and the Rights of Machines. *IEEE Technology and Society Magazine* 35, 2: 46–53. https://doi.org/10.1109/MTS.2016.2554421

[82] Jennifer Ellen Robertson. 2018. *Robo sapiens japanicus: robots, gender, family, and the Japanese nation*. University of California press, Oakland, California.

[83] Marta Romeo, Ilaria Torre, Sébastien Le Maguer, Alexander Sleat, Angelo Cangelosi, and Iolanda Leite. 2025. The Effect of Voice and Repair Strategy on Trust Formation and Repair in Human-Robot Interaction. *ACM Transactions on Human-Robot Interaction* 14, 2: 1–22. https://doi.org/10.1145/3711938

[84] Stephanie Rosenthal, Manuela Veloso, and Anind K. Dey. 2012. Is Someone in this Office Available to Help Me?: Proactively Seeking Help from Spatially-Situated Humans. *Journal of Intelligent & Robotic Systems* 66, 1–2: 205–221. https://doi.org/10.1007/s10846-011-9610-4

[85] Alessandra Rossi, Kerstin Dautenhahn, Kheng Lee Koay, and Michael L. Walters. 2017. Human Perceptions of the Severity of Domestic Robot Errors. In *Social Robotics*, Abderrahmane Kheddar, Eiichi Yoshida, Shuzhi Sam Ge, Kenji Suzuki, John-John Cabibihan, Friederike Eyssel and Hongsheng He (eds.). Springer International Publishing, Cham, 647–656. https://doi.org/10.1007/978-3-319-70022-9_64

[86] S. Šabanović, M.P. Michalowski, and R. Simmons. 2006. Robots in the wild: observing human-robot social interaction outside the lab. In *9th IEEE International Workshop on Advanced Motion Control, 2006.*, 596–601. https://doi.org/10.1109/AMC.2006.1631758

[87] Selma Šabanović. 2014. Inventing Japan's 'robotics culture': The repeated assembly of science, technology, and culture in social robotics. *Social Studies of Science* 44, 3: 342–367. https://doi.org/10.1177/0306312713509704

[88] Henrik Skaug Sætra. 2023. Machiavelli for robots: Strategic robot failure, deception, and trust. In *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 1381–1388. https://doi.org/10.1109/RO-MAN57019.2023.10309455

[89] Maha Salem, Gabriella Lakatos, Farshid Amirabdollahian, and Kerstin Dautenhahn. 2015. Would You Trust a (Faulty) Robot?: Effects of Error, Task Type and Personality on Human-Robot Cooperation and Trust. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, 141–148. https://doi.org/10.1145/2696454.2696497

[90] Andrea Saltelli, Lorenzo Benini, Silvio Funtowicz, Mario Giampietro, Matthias Kaiser, Erik Reinert, and Jeroen P. Van Der Sluijs. 2020. The technique is never neutral. How methodological choices condition the generation of narratives for sustainability. *Environmental Science & Policy* 106: 87–98. https://doi.org/10.1016/j.envsci.2020.01.008

[91] Ari Schlesinger, W. Keith Edwards, and Rebecca E. Grinter. 2017. Intersectional HCI: Engaging Identity through Gender, Race, and Class. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 5412–5427. https://doi.org/10.1145/3025453.3025766

[92] Phoebe Sengers, Kirsten Boehner, Shay David, and Joseph "Jofish" Kaye. 2005. Reflective design. In *Proceedings of the 4th decennial conference on Critical computing: between sense and sensibility*, 49–58. https://doi.org/10.1145/1094562.1094569

[93] Sofia Serholt, Sara Ljungblad, and Niamh Ní Bhroin. 2022. Introduction: special issue—critical robotics research. *AI & SOCIETY* 37, 2: 417–423. https://doi.org/10.1007/s00146-021-01224-x

[94] Sofia Serholt, Lena Pareto, Sara Ekström, and Sara Ljungblad. 2020. Trouble and Repair in Child–Robot Interaction: A Study of Complex Interactions With a Robot Tutee in a Primary School Classroom. *Frontiers in Robotics and AI* 7: 46. https://doi.org/10.3389/frobt.2020.00046

[95] Jacob Sheahan, David Chatting, Robert Collins, Jessica Bley, Alexander Eriksson, Nick Taylor, and Marco C. Rozendaal. 2024. Designing with Friction: Inverting Notions of Seamless Technology. In *Adjunct Proceedings of the 2024 Nordic Conference on Human-Computer Interaction*, 1–4. https://doi.org/10.1145/3677045.3685504

[96] Samantha Shorey, Sarah Fox, and Kristin Dew. 2017. Glimmers and half-built projects. *Interactions* 24, 6: 78–81. https://doi.org/10.1145/3140567

[97] Robert Soden, David Ribes, Seyram Avle, and Will Sutherland. 2021. Time for Historicism in CSCW: An Invitation. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2: 1–18. https://doi.org/10.1145/3479603

[98] Katherine W. Song, Samar Sabie, Steven J. Jackson, Kristina Lindström, Eric Paulos, Åsa Ståhl, and Ron Wakkary. 2024. Unmaking and HCI: Techniques, Technologies, Materials, and Philosophies beyond Making. *ACM Transactions*

on Computer-Human Interaction 31, 6: 1–6. https://doi.org/10.1145/3689047

[99] Micol Spitale, Maria Teresa Parreira, Maia Stiber, Minja Axelsson, Neval Kara, Garima Kankariya, Chien-Ming Huang, Malte Jung, Wendy Ju, and Hatice Gunes. 2024. ERR@HRI 2024 Challenge: Multimodal Detection of Errors and Failures in Human-Robot Interactions. In *International Conference on Multimodel Interaction*, 652–656. https://doi.org/10.1145/3678957.3689030

[100] Vasant Srinivasan and Leila Takayama. 2016. Help Me Please: Robot Politeness Strategies for Soliciting Help From Humans. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 4945–4955. https://doi.org/10.1145/2858036.2858217

[101] Anna Ståhl and Jakob Tholander. 2019. A Successful Failure or a Failed Success? In *Proceedings of the Halfway to the Future Symposium 2019*, 1–9. https://doi.org/10.1145/3363384.3363391

[102] Luke Stark and Jesse Hoey. 2021. The Ethics of Emotion in Artificial Intelligence Systems. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 782–793. https://doi.org/10.1145/3442188.3445939

[103] Luke Stark and Jevan Hutson. 2021. Physiognomic Artificial Intelligence. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3927300

[104] Maia Stiber. 2024. Flexible Robot Error Detection Using Natural Human Responses for Effective HRI. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, 148–150. https://doi.org/10.1145/3610978.3638365

[105] Lucy Suchman. 2002. Located accountabilities in technology production. *Scandinavian Journal of Information Systems* 14, 2: 91–105.

[106] Lucy Suchman. 2007. Feminist STS and the Sciences of the Artificial. In *The handbook of science and technology studies*, Edward J. Hackett, Olga Amsterdamska, Michael E. Lynch and Judy Wajcman (eds.). The MIT Press, 139–163.

[107] Lucy Suchman. 2019. Demystifying the Intelligent Machine. In *Cyborg Futures*, Teresa Heffernan (ed.). Springer International Publishing, Cham, 35–61. https://doi.org/10.1007/978-3-030-21836-2_3

[108] Ramtin Tabatabaei, Vassilis Kostakos, and Wafa Johal. 2025. Gazing at Failure: Investigating Human Gaze in Response to Robot Failure in Collaborative Tasks. In *2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 939–948. https://doi.org/10.1109/HRI61500.2025.10973935

[109] Leimin Tian, Pamela Carreno-Medrano, Aimee Allen, Shanti Sumartojo, Michael Mintrom, Enrique Coronado Zuniga, Gentiane Venture, Elizabeth Croft, and Dana Kulic. 2021. Redesigning Human-Robot Interaction in Response to Robot Failures: a Participatory Design Methodology. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–8. https://doi.org/10.1145/3411763.3443440

[110] Leimin Tian and Sharon Oviatt. 2021. A Taxonomy of Social Errors in Human-Robot Interaction. *ACM Transactions on Human-Robot Interaction* 10, 2: 1–32. https://doi.org/10.1145/3439720

[111] Suzanne Tolmeijer, Astrid Weiss, Marc Hanheide, Felix Lindner, Thomas M. Powers, Clare Dixon, and Myrthe L. Tielman. 2020. Taxonomy of Trust-Relevant Failures and Mitigation Strategies. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 3–12. https://doi.org/10.1145/3319502.3374793

[112] Sophie Toupin. 2024. Shaping feminist artificial intelligence. *New Media & Society* 26, 1: 580–595. https://doi.org/10.1177/14614448221150776

[113] Pat Treusch. 2017. The Art of Failure in Robotics: Queering the (Un)Making of Success and Failure in the Companion Robot Laboratory. *Catalyst: Feminism, Theory, Technoscience* 3, 2: 1–27. https://doi.org/10.28968/cftt.v3i2.28846

[114] Evert Van Beek, Elisa Giaccardi, Stella Boess, and Alessandro Bozzon. 2025. The everyday enactment of interfaces: a study of crises and conflicts in the more-than-human home. *Human–Computer Interaction* 40, 1–4: 221–248. https://doi.org/10.1080/07370024.2023.2283536

[115] Sanne Van Waveren, Elizabeth J. Carter, and Iolanda Leite. 2019. Take One For the Team: The Effects of Error Severity in Collaborative Tasks with Social Robots. In *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*, 151–158. https://doi.org/10.1145/3308532.3329475

[116] Louise Veling and Conor McGinn. 2021. Qualitative Research in HRI: A Review and Taxonomy. *International Journal of Social Robotics* 13, 7: 1689–1709. https://doi.org/10.1007/s12369-020-00723-z

[117] John Vines. 2009. The failure of designers thinking about how we think: the problem of human-computer interaction. *transtechnology research*. Retrieved August 13, 2025 from https://www.trans-techresearch.net/wp-content/uploads/2015/05/TTReader2010_010-john-vines.pdf

[118] Kelly B. Wagman and Lisa Parks. 2021. Beyond the Command: Feminist STS Research and Critical Issues for the Design of Social Machines. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1: 1–20. https://doi.org/10.1145/3449175

[119] Ronald D. White. 2022. Kicks, pranks, dog pee: The hard life of food delivery robots. *Los Angeles Times*. Retrieved September 3, 2025 from https://www.latimes.com/business/story/2022-03-17/starship-coco-kiwibot-food-delivery-bots-obstacles

[120] Meredith Whittaker. 2021. The steep cost of capture. *Interactions* 28, 6: 50–55. https://doi.org/10.1145/3488666

[121] David Gray Widder, Sireesh Gururaja, and Lucy Suchman. 2024. Basic Research, Lethal Effects: Military AI Research Funding as Enlistment. https://doi.org/10.48550/arXiv.2411.17840

[122] Katie Winkle, Erik Lagerstedt, Ilaria Torre, and Anna Offenwanger. 2023. 15 Years of (Who)man Robot Interaction: Reviewing the H in Human-Robot Interaction. *ACM Transactions on Human-Robot Interaction* 12, 3: 1–28. https://doi.org/10.1145/3571718

[123] Katie Winkle, Donald McMillan, Maria Arnelid, Katherine Harrison, Madeline Balaam, Ericka Johnson, and Iolanda Leite. 2023. Feminist Human-Robot Interaction: Disentangling Power, Principles and Practice for Better, More Ethical HRI. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, 72–82. https://doi.org/10.1145/3568162.3576973

[124] James Wright. 2023. *Robots won't save Japan: an ethnography of eldercare automation*. ILR Press, Ithaca (N.Y.).

[125] Anon Ymous, Katta Spiel, Os Keyes, Rua M. Williams, Judith Good, Eva Hornecker, and Cynthia L. Bennett. 2020. "I am just terrified of my future" Epistemic Violence in Disability Related Technology Research. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–16. https://doi.org/10.1145/3334480.3381828

[126] Xinyan Yu, Marius Hoggenmüller, Tram Thi Minh Tran, Yiyuan Wang, and Martin Tomitsch. 2024. Understanding the Interaction between Delivery Robots and Other Road and Sidewalk Users: A Study of User-generated Online Videos. *ACM Transactions on Human-Robot Interaction* 13, 4: 1–32. https://doi.org/10.1145/3677615

[127] Xinyi Zhang, Sun Kyong Lee, Whani Kim, and Sowon Hahn. 2023. "Sorry, it was my fault": Repairing trust in human-robot interactions. *International Journal of Human-Computer Studies* 175: 103031. https://doi.org/10.1016/j.ijhcs.2023.103031

[128] 2024. Students Fight Back After College Campus Is "Taken Over" By Delivery Robots. *Bored Panda*. Retrieved September 3, 2025 from https://www.boredpanda.com/starship-robots-terrorizing-east-carolina-university-students/

[129] 'Do Not Open Robots': Students Warned of Food Delivery Robot Bomb Threat at Oregon State University. Retrieved September 3, 2025 from https://www.vice.com/en/article/do-not-open-robots-students-warned-of-food-delivery-robot-bomb-threat-at-oregon-state-university/

[130] These Little Campus Celebrities are Changing College Life. *Starship Technologies: Autonomous robot delivery - The future of delivery - today!* Retrieved September 3, 2025 from https://www.starship.xyz/news/news2/