



Feminist Human-Robot Interaction: Disentangling Power, Principles and Practice for Better, More Ethical HRI

Katie Winkle
katie.winkle@it.uu.se
Uppsala University
Stockholm, Sweden

Donald McMillan
donald.mcmillan@dsv.su.se
Stockholm University
Stockholm, Sweden

Maria Arnelid
maria.arnelid@liu.se
Linköping University
Linköping, Sweden

Madeline Balaam
balaam@kth.se
KTH Royal Institute of Technology
Stockholm, Sweden

Katherine Harrison
katherine.harrison@liu.se
Linköping University
Linköping, Sweden

Ericka Johnson
ericka.johnson@liu.se
Linköping University
Linköping, Sweden

Iolanda Leite
iolanda@kth.se
KTH Royal Institute of Technology
Stockholm, Sweden

ABSTRACT

Human-Robot Interaction (HRI) is inherently a human-centric field of technology. The role of feminist theories in related fields (e.g. Human-Computer Interaction, Data Science) are taken as a starting point to present a vision for Feminist HRI which can support better, more ethical HRI practice everyday, *as well as* a more activist research and design stance. We first define feminist design for an HRI audience and use a set of feminist principles from neighboring fields to examine existent HRI literature, showing the progress that has been made already alongside some additional potential ways forward. Following this we identify a set of reflexive questions to be posed throughout the HRI design, research and development pipeline, encouraging a sensitivity to power and to individuals' goals and values. Importantly, we do not look to present a definitive, fixed notion of Feminist HRI, but rather demonstrate the ways in which bringing feminist principles to our field can lead to better, more ethical HRI, and to discuss how we, the HRI community, might do this in practice.

CCS CONCEPTS

• Human-centered computing;

KEYWORDS

feminism, research methodology, design methodology

ACM Reference Format:

Katie Winkle, Donald McMillan, Maria Arnelid, Madeline Balaam, Katherine Harrison, 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 (HRI '23), March 13–16, 2023, Stockholm, Sweden. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3568162.3576973>

1 INTRODUCTION

Feminist Human-Computer Interaction (HCI) has been an established concept since (at least) 2010 [10] and, more recently, Critical Data Studies has been engaged in a conversation about the way feminism could make data science better, and more just [28]. In their recent book, *Smart Wives*, Strengers and Kennedy call for a 'feminist reboot' of assistive technologies (including robots) to ensure more ethical designs [104] – but what could this 'look like', in practice, for Human-Robot Interaction (HRI)? We note that, compared to HCI for example, there is one aspect of HRI which makes feminist theory even more relevant for us specifically: roboticists generally work with embodied technology. We are, often, actually building or designing (robotic) bodies and identities. Feminist theory, too, has dealt extensively with embodiment, delving deeply into the relation between the material body, the social, and the subject [16, 24, 43], not least in relation to technology [47, 78, 85, 94, 98, 106]. For this reason, we posit that feminism has several epistemological approaches and methodological insights that can help us in our work. To show this, we first explain what we mean by 'feminist', following this with five principles for Feminist HRI, drawn from Feminist HCI and Data Feminism but discussed with reference to HRI literature and applications. We then reflect on HRI practice, presenting ten questions designed to encourage awareness and change throughout the design and development process. We conclude that an awareness of feminist principles when designing *any* HRI will result in better, more ethical robot deployments overall.

1.1 What Do We Mean by Feminist?

Feminism is sometimes a charged and emotive word, and can be used in many different ways to mean many different, and at times conflicting, things. Here, we use the term to indicate an approach to the design and development of robots and human-robot interactions which is sensitive to power structures. Important in



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

HRI '23, March 13–16, 2023, Stockholm, Sweden
© 2023 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-9964-7/23/03.
<https://doi.org/10.1145/3568162.3576973>

such an approach is an attention to the intersections of different power structures, based on gender, race, class, ability, sexuality, religion, and so on. This concept of intersectionality, a widely accepted concept today that originates from Black feminist schools of thought [19] through the work of Black feminist scholars and activists dealing with interlocking and overlapping systems of oppression [22, 23, 51, 74]. We posit that a feminist perspective can be used to analyse and inform the robots/HRI we are designing, as well as the way we actually ‘do’ HRI research in terms of methodologies and community practices. A feminist approach informs not only what work is done and how, but also calls on researchers to be accountable for their own positionality, privilege and point of view.

We suggest that the principles of Data Feminism, “*a way of thinking about data, their analysis, and their display, that is informed by [...] feminist activism as well as the legacy of feminist critical thought*” [28] in conjunction with Feminist HCI interaction qualities [10] provide a comprehensive and actionable starting set of principles for developing the concept of Feminist HRI. The first two principles of Data Feminism are: #1 *examine power* and #2 *challenge power*. In Section 2, we co-opt these principles in pitching them as two key steps or stages of Feminist HRI, grouping together the other Data Feminism principles alongside Bardzell’s feminist interaction qualities as ways of achieving these high level goals in practice. We present these principles, quoting directly from these two source texts, under each subheading of Section 2, wherein we discuss how they might map to HRI. The overlap and shared methodologies between the fields of HRI and HCI [53, 82] make Feminist HCI an obvious place to look for inspiration, complemented then by the broader research/development pipeline consideration, and centering of sensitivity to power, offered by the Data Feminism principles. Other pertinent paradigms being discussed in ongoing feminist-technology discourse include Value Sensitive Design (VSD) [14, 112], the Ethical Computing Platform [77] and Critical Race Theory [76]. Additional authors who have written foundational texts about how identity and ‘the social’ impacts science and technology research in general include [9, 44, 45, 64, 110]. Studies of how power and social justice are entangled with engineering and robotics research can be found in [5, 33, 46, 47, 83, 86, 99, 106, 115]. We note also that the interdisciplinarity of robotics means that works concerning e.g. biases in artificial intelligence (AI) [11, 15, 59] and educating the general public about these biases [67, 92] also provide insightful examples of (what we would call) feminist approaches to technology that HRI researchers might look to draw upon.

1.2 What Do We Mean by Power?

D’Ignazio and Klein’s concept of power is as follows: “*We use the term power to describe the current configuration of structural privilege and structural oppression, in which some groups experience unearned advantages – because various systems have been designed by people like them and work for people like them – and other groups experience systematic disadvantages – because those same systems were not designed by them or with people like them in mind*” [28, p. 24].

Identifying robots as embedded within subject-positioning relations¹, as (robot) bodies which simultaneously reflect and influence

structures of power, forces us to think about embodiment as a practice, rather than a material artefact. This requires us to think about robots and their design, people and bodies as positioned differently in and by structures of power rather than merely possessing demographic characteristics [18] as we typically tend to do when talking about (social) robot design (e.g. [81, 118]) and robot users (c.f. Friederike Eyssel’s HRI 2022 keynote talk on the importance of sex and gender analysis [32]). An example can be seen in critique surrounding Shudu Gram, a computer generated model presented, for all intents and purposes, as a black woman model/“influencer” [55]. Feminist discourse on Shudu may indeed talk about risks and opportunities regarding the specific choice of ‘her’ race and gender, but would likely be more concerned with Shudu actually ‘belonging’ to (i.e., being monetised by) an Australian white man². Recognising Shudu and ‘her’ creator as positioned within structures of power, i.e. in the context of a modelling industry rife with gender bias, racial biases and abuses of power [70], highlights why issues of ‘her’ aesthetic design are more than ‘skin deep’.

In short, it is important that Feminist HRI does not become synonymous *only* with questions of e.g. norm-breaking robot design cues pertaining to gender [118] – in the words of Katta Spiel, in their recent keynote talk at the 2022 RO-MAN workshop on Gendering Robots: “*queerbots won’t save us*” [1]. Instead, Feminist HRI implies considering the contexts that robots are used in and how people around them (users, procurers, assistants) are positioning the robots in material-discursive practices, rather than focusing on the shape, color or clothing of the robots we use. This first requires, per Data Feminism principle #1, *examining power* [28, ch. 1].

2 FEMINIST PRINCIPLES FOR HRI

2.1 Step 1: Examining Power

Begin by analysing how power operates in the world [28, ch. 1].

In an internal workshop, we attempted to sketch out some of the subject-positioning and power-relations around key actors (people, robots, resources etc. [68]) involved in delivering HRI, from initial funding of an HRI research project through to deployment of the/a resultant robot into society. The result is shown in Figure 1.

Figure 1 might seem rather obvious to some readers. Sabanovic’ mutual shaping framework, published over a decade ago, already called for roboticists to recognise the two-way interaction between robotics and society [89]. Her work challenges the techno-deterministic narrative, and has led to increased recognition of participatory design and ethnographic studies highlighting the importance of, and benefit from, trying account for the broader social context a robot will be embedded in – c.f. the left hand side of our figure. One thing the feminist perspective brings, via sensitivity to power, is the recognition that these mutual shaping interactions are not necessarily equal or just. For example, in Figure 1 we have drawn the reverse arrow back from the user to the robot as dashed rather than solid: whilst the user no doubt shapes and may even co-opt robot deployment to their needs, they likely have limited control over key variables of robot deployment including at the very highest level, whether the robot is deployed (at all) and if/when it is removed. Engstrom and Jebari [31]’s damning account of the negative impact

¹“the concept of subject positions holds that people make sense of who they are (and are understood by others) within culturally circulating discourses and narratives”

²Discussed in Ketchum’s Intro to Feminist and Social Justice Studies Podcast [58]

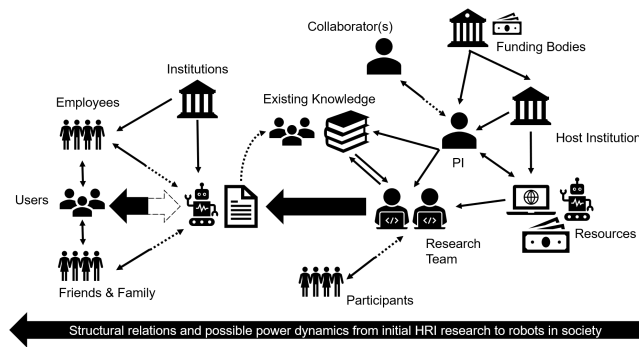


Figure 1: One possible sketch of relations on thinking about how HRI research(ers) and the robots we produce are positioned with structures of power. Dashed arrows indicate potential (but unequal power) for mutual shaping [89].

of robot deployment on workers in Amazon fulfillment centers provides a concrete example of such unequal relations.

Examining power is a first step to identifying what we can do, within our sphere of influence, to account for (and even challenge, per Section 2.2) such dynamics. This brings us to the right hand side of Figure 1, which is concerned more with power structures at play during the HRI research and development process, as a Feminist HRI approach calls also for researchers recognising their own power and positionality³. Variation in such power is also something we have tried to consider when identifying ways to practice Feminist HRI. A PhD student likely has limited power compared to a principal investigator (PI). For example, they might have no say in choice of robot platform, but perhaps they do have power to cite underrepresented researchers alongside those classic texts which seemingly get cited ‘by default’ [4], or to actively implement best practice guidelines regarding collection and reporting of participant gender data [90]. The PI instead might reflect on who they are (not) collaborating with, whilst we as a community might look to recognise that PI access to resources can vary significantly by institution and geographical location. Robots and human-subject studies are expensive – the double-blind review process of the HRI conference currently means that these factors cannot be taken directly into account, so perhaps when reviewing HRI work we should avoid default skepticism of works which utilise virtual robots, video-based studies or small numbers of participants, if the work provides a novel, timely contribution and identifies/reflects on any resultant methodological limitations.

2.1.1 Considering Context, Ecology: “Considering impact of deployment on the broadest possible context and widest possible range of stakeholders.” [10]; “asserting that [HRI] is not neutral or objective, rather the product of unequal social relations; context which is essential for conducting accurate, ethical analysis” [28, ch. 6].

Examination of power predicates consideration of the broader context a robot will be deployed into. As alluded to above, there exists already a significant body of work which considers the broader social context around robot deployment. For example, Chang and

³c.f. recent discourse on the implicit tension entailed in trying to do social justice work within the confounds of being an HCI researcher, or working within the academy [61]

Sabanovic used a social shaping framework to note social factors impacting on/affected by long term robot use in a care home [17]. Forlizzi’s ethnographic study of robot versus conventional vacuum cleaners is another great example of this, as it discusses impact of deployment on the household cleaning regime and related social dynamics, over-time, rather than focussing solely on an individual, snapshot 1:1 user-robot interaction as typically assessed in HRI user studies [36]. Other works more explicitly reflect on the potential impact of HRI on those observing interactions between robots and their primary user. An example is Garcia Goo’s work [2] on the impacts of observing robot abuse, in which the authors build on prior experiments on this topic [87] to take a more intersectional approach in understanding how such impacts might vary across (robot and) participant gender, participants’ prior experience of victimisation and levels of sexism.

For help in identifying pertinent power dynamics within their application domain, HRI researchers can also look to the application-specific feminist commentary. For example, much HRI research is concerned with robot-supported healthcare [35, 57, 80]. Feminist scholars have identified and discussed at length the complex power and control dynamics which might exist between care receivers and their (in)formal care givers [27, 35, 57, 80]. HRI researchers might then reflect on whose interests the robots and interactions they are creating really serve. Will they increase autonomy and independence of care receivers, or rather further increase power imbalances between care receivers, (in)formal care givers and/or the institutions charged with delivering care? Asserting that HRI is not neutral forces us to reckon with these questions and recognise the ways both the research we do (e.g. in the form of whose data across what metrics we deem ‘valuable’) and the robots/interactions we hence create or motivate might reinforce systemic structures of oppression.

2.2 Step 2: Challenging Power

“commit to challenging unequal power structures and working towards justice” [28, ch. 2]; “Avoiding both the propagation of harmful stereotypes, but also the imposing of designers’ own values on users/stakeholders” [10].

Examining and reflecting on power per Section 2.1 already offers a route towards more ethical HRI design and development. For those that want to go further, a next step in Feminist HRI would be to *challenge* it. Again looking to HCI as our nearest neighbour, ideas of advocacy and social justice feature strongly in recent Feminist HCI works, moving from the continued stream of work that is invested in examining and critiquing inequalities in technology and design towards using technology and design to highlight and respond to these inequalities. In this body of work we start to see both the role of technology in making change, and also the work that must happen around technology – at a social, political and structural level – for change to occur. Such work demonstrates that HCI research/practice can legitimately happen at each, and all of these levels (technological, social, cultural, political, infrastructural). For example, Ng et al. [75]’s recent examination of the necessary political and grassroots work required of an activist and entrepreneur to introduce the menstrual cup into Taiwanese society shows the complex web of barriers that must be navigated to

make lasting change in body autonomy. Balaam et al. [8]’s use of a mobile phone application to enable a breastfeeding community to collect large-scale data on the social barriers to breastfeeding in public highlights how simple tools can be used by communities to highlight matters of concern with a potential to make local change, and Strohmayr et al. [105]’s use of digital technologies illustrates how technologies should be considered more than as a tool for interaction, but as a catalyst towards change for a better world.

In describing the quality of advocacy, Bardzell points to a difficult intersection for (feminist) designers who seek to ‘*bring about generate political emancipation and not just keep up with it*’. Whilst the term might sound provocative, we can interpret *political emancipation* here to mean the empowerment of (typically marginalized) people to engage in something previously out of reach. If we put very typical HRI applications on assistive technologies aiming to increase independence and dignity e.g. persons with disabilities [12, 113] at one ‘non-provocative’ end of this spectrum, then we might put (arguably) more provocative work looking to use robots in encouraging moral behaviour and/or challenging societal stereotypes [60, 63, 118] on the other, with group-robot applications that look to equalise participation and improve inclusion somewhere in-between [40, 41, 109].

Using empirical research to inform work on such applications leads to focus on/working within the status quo, and hence risks propagating regressive and/or harmful norms and practices in the name of trying to build “universally effective” systems (the fallacy of *universality* being something else a feminist approach must recognise – discussed more under Section 2.2.2 on Pluralism). However, attempting to create ‘more progressive’ design outputs runs the risk of designers imposing their own values and ideas about what an ‘improved society’ looks like. Bardzell posits participatory approaches as being one clear way to tackle this, as they distribute authority and responsibility for decision making across stakeholders – thus the aforementioned overlap with participatory design.

2.2.1 Participation, Making Labour Visible: “*Valuing participatory processes that lead to the creation and evaluation of prototypes*” [10]; “*making visible the many ‘hands’ involved in undertaking [HRI]*” [28, ch. 7].

Participatory design is relatively well established within HRI such that we can already point to excellent examples of works embodying the participation principle. One such work, which arguably engages participation within an advocacy context, is Valencia et al. [113]’s work on co-designing socially assistive sidekicks as augmenting and alternative communication (AAC) devices for improving communicator inclusion in conversation. By including both AAC users and their conversation partners, the work also showcased the principle of *considering context* per Section 2.1.1. When working with marginalized communities, we can also look to Feminist HCI for insight on how to ensure these discussions are attended to in ways which are meaningful and ethical (e.g. [48, 101]), efforts also recently mirrored by HRI researchers [6, 124] to complement previous suggestion of a code of ethics for the HRI profession [84].

The concept of making labour visible builds on that of participation in two key ways. Firstly, it is about properly valuing, crediting, and reimbursing all of those contributions required to ultimately bring an HRI development or deployment to fruition. As pointed out

in Data Feminism, improper remuneration of research participants (including a reliance on voluntary participation) disproportionately excludes women, the working class and people of colour from taking part in research. A recent review indicates this may indeed have been an issue in HRI research to date, providing recommendations around recruitment to address (primarily gender) representation in HRI research [117]. Secondly, it is about better crediting and elevating different forms of labour. A tangible example highlighted in Data Feminism is the Atlas of Caregiving, whose pilot study [66] utilised HCI devices in conjunction with qualitative data collection to generate a comprehensive documentation of the various tasks involved in family caregiving. We struggled to identify existing HRI works which provide clear exemplars of this (outside of related works previously discussed around e.g. participation, ethnography and later under emotion and embodiment), but note it is also something called for in Strengers and Kennedy [104]’s call for a feminist reboot of social and assistive technologies. They discuss how such technologies can support (men’s) increased engagement in traditionally feminised labours, but identify a need to ensure we raise the status of such labours for *everyone*, not just make them cool and masculine via ‘mentrification’ with gadgets and toys [104, ch. 8, p215 “*New Toys for Boys*”]. We suggest one possible starting point is to at least question when robot autonomy is appropriate and recognise instead when teleoperation might be better e.g. in socially assistive robots [30] thus recognising the skill and value in caregiving work rather than assuming it is something (a) easy and (b) desirable to automate.

2.2.2 Pluralism: “*Resisting any single, totalising or universal point of view*” [10]; “*a practice of “synthesizing multiple perspectives”*”; “*insisting that the most complete knowledge comes from synthesising multiple perspectives, prioritising local, Indigenous and experiential ways of knowing*” [28, intro; ch. 5].

HRI works concerned with examining cultural differences in robot evaluations, and/or looking to draw from different cultures in robot design, contribute to the quality of pluralism. For example, works discussing Confucian ethics perspectives on robot blaming [122] or robots as moral advisors [60] challenge the notion of ‘western-by-default’ cultural contexts when thinking about the expectations users have from robots, and the way to approach designing behaviour for morally-laden scenarios. Works which include multiple experiment sites to allow for cross-cultural comparisons [103] are also important, but we should be careful to avoid a model of simply ‘exporting’ studies done in the west to replication sites elsewhere in order to ‘check for generalisability’, and rather start from an understanding and recognition that we should not expect *universal* HRI experiences. Pluralism can also pertain to *considering context* and *participation*, for example, Tanqueray et al. [108] invited gender studies scholars and clinical staff to (independently) discuss an application of socially assistive robots in women’s reproductive health – noting that each group had significantly differing views on the risks and opportunities such an application poses. We were unable to identify any HRI works which similarly spotlighted local, Indigenous or experiential knowledge, and so we point to examples in HCI from which we might take inspiration [3, 7, 71, 114] and identify this as a significant research gap within the HRI literature.

2.2.3 Elevate Emotion and (User) Embodiment. “Recognising that users have (gendered) bodies which experience sexuality, pleasure, desire and emotion.” [10]; “value different forms of knowledge, including that which comes from people as living, feeling bodies” [28, ch. 3].

Whilst robot users’ bodies have been identified as the source of multimodal cues, some of which might be associated e.g. with personality or affective state of the user for informing or evaluating robot behaviours [39, 42, 62, 65, 93] very few works are seemingly driven by a desire to generate some bodily experience in users, or even simply *ask* participants how they felt during an interaction. This is evidenced in a recent survey of metrics and methods utilised in over 1400 HRI research works published at the fields two main conferences from 2015-2021 [123]. The authors identified and categorised the most frequently used subjective and objective measures, with the results pointing towards a lack of consideration regarding users’ bodily experiences, in place of focus on opinions about the robot or more objective measures pertaining primarily to ‘task efficiency’. An exception is the documenting of users’ emotion directed at the robot, seen in 216 papers. This represents most common type of subjective measure, but still less than fifty percent of the 463 papers seen to be measuring the most common objective measure, which was ‘task performance’. Further, attention should also be paid to the motivations underpinning those works which do examine user emotion expression – is there genuine interest in embodied user experience, or rather a desire to e.g. evidenced improved robot acceptance via a proxy quantification of bodily signals?

We can take further inspiration from the breadth of ways Feminist HCI has looked to account for the (diverse) human body in HCI at large. This includes the application of approaches such as soma design [50] to interrogate and design for new types of experiences and interactions with technology which account for the body and its processes (e.g. [97, 102]) through to critical alternatives to the quantification and commodification of the human body so commonly seen in industry [49]. Treusch et al. [111]’s work using robot knitting to question the narrative of optimisation provides perhaps a first example specifically utilising HRI this way. Centering bodily experiences like this, in place of the typical ‘task efficiency’ measures noted above, represents a challenge to power as it shifts the narrative regarding what success ‘looks like’; altering how our research findings might be utilised in future works/commercial developments and hence who our research outputs serve. Aiming to maximise measures of user comfort in collaborative HRI settings (perhaps combining subjective user-reports with objective measures pertaining to e.g. heart rate, electrodermal activity, skin temperature etc. [120]) rather than minimising task speed, might help prevent HRI research fuelling situations like that in the Amazon warehouse, where the deployment of robots centers speed and efficiency over all else, with detrimental impact on human workers [31]. Again here it is an issue of *which* measures and *why*, moving away from an unquestioned assumption of needing to improve acceptance, demonstrated via some quantified gain in a particular bodily signal measure, but rather on working towards a better experience for the user.

Works on affective touch in HRI, typically trying to provide those unable to experience (human or animal) social touch, demonstrate the embodiment quality in their focus on the (bodily) benefits of

such; be it e.g. reduced stress and release of oxytocin [13] or reduced pain [56]. Block et al. [13]’s work on robot hugging specifically also showcases the embodiment quality in the way they conceptualise their design goals/assess interactions as they refer to designing for ‘pleasant, natural-feeling hugs’ which users’ are asked to evaluate via scales from ‘love’ to ‘hate’, reflecting those more primordial and experiential user feelings Bardzell refers to. Building on our comment above regarding motivations for capturing user emotion expression, these affective touch works can be contrasted directly with works using robot touch as a persuasive cue, primarily considering impacts of this on some desirable user behaviour such as task engagement [73], self-disclosure [96] or pro-sociality [95]. More generally, works which utilise measures pertaining to examining users’ bodily or affective experiences of HRI, ideally as a goal in their own right rather than being proxies for some other more typical measures of engagement or perception of robots, might be said to be exhibiting the quality of embodiment. The physical robogame Robotower, and de Oliveira et al. [26]’s examination of how to maximise user *fun* offers an example.

We suggest Fosch-Villaronga and Poulsen [37] provide an example of (re-)considering how the embodiment quality might manifest in healthcare robotics with their discourse on the potential use of sex robots in care for persons with disabilities. They identify that sex robot applications might be sexual, emotional or educational in this context, where the sexual theme specifically includes e.g. the intimate act of cuddling, something which significantly predicts sexual satisfaction of men with disabilities, but is unlikely to be what people first think of when imagining sex robot applications. In doing so, the authors look to challenge the default narrative of sex robots, legitimising and making space for more nuanced discussions around treating users as whole bodies when conceptualising robots for healthcare. This work, along with Dudek and Young [29]’s suggestion that sex robots taking more inspiration from the 2LGBTQIA+⁴ community (itself hence demonstrating also the principle of *pluralism* discussed previously) could offer exciting potential for users to explore their sexual identity, are seemingly the only works considering positive bodily experiences that might be derived from HRI with sex robots currently found in the proceedings of the International Conference on Human Robot Interaction. Other works discussing sex robots are concerned with user perceptions of sex robots and their potential to (negatively) influence e.g. human-human interactions [52, 88, 91]. A feminist perspective identifies new points of consideration on either side of this discussion. For example, it motivates design exploration of sex robots designed according to broader experiences of sexual pleasure in place of the focus on male pleasure and penetrative sex seen to date [104]. On the other hand, it would also call for careful consideration of the implications regarding autonomy, consent and coercion when replacing sexual intimacy with a humanoid object and if/how that may change a humans perspective on consent and autonomy, paying particular attention to potential harms for people of marginalized identities within patriarchal societies.

⁴“an inclusive acronym to refer to the noncisshet (cisgender heterosexual) community [...] including Two-Spirit, Lesbian, Gay, Bisexual, Transgender, Queer/Questioning, Intersex, Asexual/Agender/Aromantic (2LGBTQIA+)” Dudek and Young [29]

2.2.4 Re-Think Binaries and Hierarchies. “challenge the gender binary along with other counting/classification systems that perpetuate oppression” [28, ch. 4].

A number of critical works within HRI have recently drawn attention both to reporting practices [20] and the counting/classification systems we are utilising more broadly (although more in the context of critiquing the lack of standardisation [123]). The review on HRI research participation [117] we previously discussed under Section 2.2.1 is also relevant here, as it identifies a practice of excluding (inherently or explicitly) non-binary research participants (who typically represent a very small number of participants) when conducting (quantitative) gender analysis, and provides a number of starting suggestions for going beyond this. In HCI, critical works have also drawn attention to the harms of designing for users within the binary definition of gender, for example Spiel [100]’s auto-ethnography on navigating technology-related gender registrations as a non-binary individual. We note that calls for increased gender and sex-based analyses in HRI [32, 107] need not be mutually exclusive, but should be explicitly articulated and carefully captured with informed consent, as we must also be aware of the risks this poses for entrenching sex/gender binaries and stereotypes [21, 38]. One obvious way HRI might look to challenge gender binaries is in the design of non-binary or gender ambiguous robots, a design space recently alluded to by multiple researchers [81, 118], although we refer again to the need to consider particular robot design choices within the subject-positioning relations as per §1.2.

2.2.5 Self-Disclosure. “Making visible the ways that technology shapes users towards become the projected ‘ideal’ user.” [10].

We struggled to identify HRI works which might be said to exemplify this principle. Bardzell cites Amazon’s recommender system as an example: users can ask which previous purchases led to a specific product recommendation they have received, and indicate whether certain purchases should be ignored for this purpose [10]. We might point to works on robot transparency/explainability [34, 119] as being somewhat relevant, although perhaps missing elaboration regarding this notion of the ‘ideal user’. An example from Feminist HCI on wearable technologies perhaps highlights this element better, as it discusses the use of biosensor-embedded clothing to ‘conceal and reveal’ users’ body data in the context of challenging narratives around cultural expectations and ownership of identity [54], although this perhaps centers *user* self-disclosure where we (and Bardzell’s recommender example) might think more about *robot* disclosure. Regardless, we identify this as one principle representing very much an open design space for the community to reflect on and co-opt to (re-)define what designing for user and/or robot self-disclosure might encompass.

3 TOWARDS FEMINIST HRI PRACTICE

Drawing from the concepts and examples outlined above, here we try to summarise some starting points for Feminist HRI practice. We focus on how these principles are made apparent in the day-to-day work of researching and designing robots interacting with humans, either in the lab or in the wider world. To make these actionable we split this section around the design of the robot and the interactions with it, and the wider design and reporting of the research itself. While this is, to an extent, a false dichotomy, it provides a structure

to our recommendations for practice – allowing those with different roles at different times the freedom to reflect on the complex interactions between them. Within each we focus on two actions to take, through a series of questions⁵ to be answered about the robot, the users, the research, and the researchers themselves. We propose that, in order to propagate and normalise taking a feminist approach to HRI, these questions need not only be seen as tools for self reflection and planning, but can be explicitly answered in design briefs, corporate reports, and research publications. Whilst in sum representing a novel way to think about HRI, we do not propose any new methodology specifically, but rather suggest these questions can be explored with current UX, HCI and/or HRI methods that practitioners use to identify opportunities and challenges for interventions, involve users in ongoing design processes, and verify the robot is fit for purpose.

3.1 The Robot

Robot appearance and ‘personality’ might be the first point of contention that comes to mind on hearing the term *Feminist Robots* (c.f. our own previous work in this direction [118]). Yet here we want to quickly move beyond looking at the presentation of the robot in isolation to dig down and reflect on the robot’s subject positioning more broadly, which is what really guides if, how and why particular design choices matter.

3.1.1 Examining. In HRI we know that in many cases the physical appearance of the device is highly constrained. The choice may be limited to a single device available at the time and place of the research, and creating custom hardware is unnecessary for answering many questions that HRI research sets out to investigate. What is also clear, from the principles above, is that it is not the physical appearance of the robot *in isolation* that is important, but rather how the embodiment, including any specific choice of anthropomorphic social identity cues, relates to its subject position within the broader social context. To start exploring this, researchers might consider:

What are the norms and expectations around the robot’s actions?

The first of the actions we propose to be taken is to examine the role that the robot is designed to take in the interaction, with a focus on what that role means in practice. By this, we encourage designers to systematically map out the what a *human* performing the same actions in the same situation would be expected to know, how they would be expected to act, and how they would expect to be treated. As part of performing this mapping, we encourage the designer to examine their own expectations around this role – who is the person that you expect to fit into this interaction, and why?

Moving outward from the robot, the next step of this process of examining the human-robot interaction through a similar lens.

Who will be interacting with the robot?

While keeping some ambiguity in mind as to the robot’s level of autonomy in this interaction, it is important to map the people that will be the subjects of these interactions. In doing so we encourage designers to expose the expectations that the user will have the first, and the thirty-first time they perform this interaction. How these expectations are shaped by the human-centric exploration above is an important set of relationships to map in understanding

⁵we take inspiration here from the Feminist Design Tool put forward by Josie Young and the Feminist Internet [121]

how the robot can effectively and efficiently fit into the use case. Beyond examining the preconceptions and learning of the users that will directly interact with the robot, it is also important to map any secondary and tertiary users, their interactions with each other and the primary user. Document what you expect an observer would see and be able to understand from the interaction, and how a user may explain or discuss their use – with an observer or someone who had never seen the robot. In doing so this gives you the opportunity to expose your understanding of how the social environment is shaped by the change the robot represents, and plan for it accordingly. The previous literature on social shaping and in-situ HRI studies highlighted earlier demonstrates how this can be done in practice.

How will the robot change the interaction?

We have not yet mentioned the proposed autonomy of the robot, focusing first directly on the human users and their actions, interactions, and expectations, generally with reference to an equivalent human-human interaction. The next stage is to explore exactly how the translation from prospective human-interaction to instantiated robot-interaction changes the material and immaterial qualities of that interaction. Currently, no human activity can be perfectly and seamlessly carried out by a robot – a limitation that looks set to continue for the foreseeable future. Instead of glossing over the important technical and design work that is behind the translation, tuning, and framing of an interaction to make it robot-compatible we suggest that it be documented and reported fully. We might consider a mechanism akin to model cards as put forward in the machine learning community [69]. In doing so, the change that the robotic interaction has on the interaction and those who will interact with it can be better understood, and the design itself can be iterated on to balance the needs of the interaction with the social and physical situation that interaction happens within. Maintaining a sensitivity to power through this process can be achieved by asking – what is being gained, by who, from these interactions? And, on the contrary, what is being lost, by who? Further, who is (now) doing what kind of labour in order to support them?

3.1.2 Challenging. The examination of the situation and the norms which the robot will interact through and with, and how they will be changed to successfully complete the goals of the intervention design allows for a more nuanced and better situated robot interaction. However, as developers and designers we have the power to shape interactions, and through them those situations in which the interactions take place. Building on the understandings developed in the *examining* phase the next step, the next challenge for those designing and developing robots, is to decide where and how to affect change.

What norms do we want to promote, and what ones to challenge?

While the examining phase exposes the norms and attitudes surrounding the interactions being designed, they should not be automatically encoded directly into the behaviour and design of the robot. Rather, designers and developers should make conscious and informed decisions regarding which norms to replicate versus which ones to challenge, taking into account the wider situation uncovered in the previous step, extrapolating the impact of such encoding on a wider range of users, including those in a secondary and tertiary role. Replication of norms may decrease the learning

phase or increase intuitiveness of interaction, or even increase acceptance in some demographics, but also risks propagating social inequalities ranging from stereotyping [104, 116] to dehumanization [31, 79]. We encourage designers to explicitly report on what roles and norms they have chosen to take advantage of to ease the interaction, which ones they have chosen to challenge through changes to the appearance or actions of the robot, and why either path was taken.

Can we minimise the risk of harm to low-power users by reducing, or at least not exaggerating, situational power imbalances?

Having an eye on the latent power dynamics within the context of use, designers are then able to consider what active influence they have in the robot deployment to alter them. One example would be designing informed, enthusiastic consent *into* ongoing robot interactions, as has been suggested in mobile devices [72], rather than trusting it to institutional actors around the robot. Important here also is our choice of data input streams and ‘success measures’ that might be driving an interaction, per our previous discussion on centering bodily experience over ‘task efficiency’. Additionally, designers can also reflect here on what labour the robot is doing and how to make visible (rather than diminish) the labour of their human collaborators.

3.2 The Research

The second phase of *examining* and *challenging* we encourage HRI practitioners to engage in involves taking one step back from the design and deployment of the robot to encompass the design and implementation of the research itself. We take this in a chronological order, starting with the development of the research idea, through the planning of the development and research trial, the metrics and data collection and analysis, through to the reporting of the results and the sharing of relevant data and artefacts. These steps cover a range of different roles and responsibilities, and we recognise that not all of those involved have the opportunity nor the power to challenge at every stage, but by contemplating ways in which to enact change we hope to encourage an openness in the system, an understanding of the different constraints placed by the competing pressures of academia, and a willingness to consider change where and when possible.

3.2.1 Choosing Challenges and Defining Research Questions. The first stage of any research project is the decisions as to what challenges to address and what research questions to answer.

Why are we answering these questions about this population?

Whilst many might point e.g. to funding initiatives as being a major force in setting research direction, it is rare that any research program or question does not have the ambiguity to allow for the researcher(s) to adjust it in some way to support their own goals and values – goals and values which are likely to evolve in response to the explorations described above. Periodically checking-in to reflect on and refine/re-evaluate initial research questions/goals, particularly in cases where researchers might be working with a specific population they’ve limited previous experience of, is a worthy exercise. The *we* within this question is also important and intentional. In finding peers to discuss research proposals, implementations, and plans for studies and publications there are decisions made on

who to include in these conversations. These can happen at different levels of the academic hierarchy and in choosing collaborators, there are any number of competing goals and constraints on geographic location, research background, reputation, and personality that are – to greater and lesser extents – taken into account. Where such collaborations are the norm, we would encourage reflection on the diversity of inclusion taking place – is it the ‘usual suspects’ from a small number of labs? Why is that the case? Where there are barriers to such collaboration, those barriers themselves should be investigated. At a community level, perhaps we can do more to facilitate networking and (unexpected) collaborations e.g. in the form of workshops (c.f. the DEI Workshop at HRI 2022[25]) or themed networking events at our conferences.

We recognise the extra labour involved in working with a variety of people with different backgrounds and views on academic practice, but see this as a well spent investment (we as an author team have certainly found so, in doing the work underlying this manuscript). A diversity of voices at all stages of the process not only results in more grounded results and better generalisability but also increases the visibility of those results and the ongoing research through multiple social and collegiate networks.

Why are we recruiting these participants?

Consider the choice of target users when defining what to implement and test, drawing from this who to recruit and how to support their taking part in the study. At each phase we encourage researchers to challenge the status-quo of samples of convenience (including e.g. with respect to need for reimbursement or office hours availability) in participant recruitment. Again, we recognise the extra labour involved in making research participation accessible to a broader range of participants, and again we argue that the resultant diversity of data gathered will return dividends in research quality, although even if it didn’t, we’d argue for it anyway because equalising access to opportunities is important for the communities we live and work in.

3.2.2 Conducting the research. When conducting research each field and each method employed has its own set of norms and procedures to be followed. What we want to bring to the fore is that these methods and procedures are not static, with methodology papers regularly published across many disciplines challenging and changing the way research is conducted from multiple fronts.

Why are we employing this methodology in this way?

We want to encourage the same level of reflection on the norms and practices of the research itself as of the situations our technologies are designed to support and change, and to do that we must be cognisant of the malleability of research as a practice. Researchers can consider the practical aspects of the method being used, the different roots in theory and practice that different versions of analytical practice have, and the impacts it has on the people and the power structures it touches.

3.2.3 Metrics and Analysis. Here there are many choices that are made along the way in the research procedure that come together in the analysis of the collected data, providing the contribution of the research as a whole. The first is the metrics that are collected during the process. Following on from the methodological questions asked above, what you chose – or chose not – to measure can have a significant impact on the types of contributions and how they

are understood by the research team. We refer readers to Data Feminism [28] for detailed further discussion on this, but point again to a question we raised earlier: when evaluating robot design – who is gaining, who is losing, and which impacts are actually being measured?

How can we allow for participant agency?

We have so far discussed only the researcher’s needs and perspective – but what about participants? We suggest that participant metrics be exposed to them when they are available. This can be after the fact, but having a procedure to allow participants to see what was collected, why, and decide with some granularity what to allow the researcher to use, what to remove, and – depending on the data – what to correct.

This call for agency follows through to the analysis of the results. There are many steps in the analysis in which the researcher makes choices that determine the focus of the contributions. These can be from which metrics to focus on, which outliers to remove, to what bins are used to split and compare participants or interactions. We suggest that in making these decisions researchers should look beyond the methodological norms and the goal of presenting a contribution to consider what such choices mean for the participants being measured, the type of contribution that the researcher wants to be making, and how it fits in the wider research conversation on this topic as a whole. Questioning why participants are split by gender, age, or geography (for example) starts with reflecting on the reasoning on why such a difference may exist in the first place. Allowing for agency here, both on the part of the researcher and the participant, means identifying what such results might enforce or challenge in the wider world, and if those involved are happy with their labour being employed to do so.

4 CONCLUSION

We have presented a set of feminist principles for HRI, drawn from our neighbouring fields of HCI and Data Science, and identified existing HRI works that exemplify these qualities already, otherwise discussing how examples from these other fields might translate into HRI. We posit that, first and foremost, pursuit of these principles will yield better, more ethical HRI through increased understanding and conscious decision making relating to subject-positioning power relations in the research/design we do and the user-robot interactions that result. For those interested in going further, explorations of how we might challenge such power relations motivates exciting new directions in HRI research. We have provided a set of reflexive questions designed to support researchers at either level, first in honing their sensitivity to power, then their ability to challenge it. In doing so, we hope to provide a starting point for further discussions regarding what Feminist HRI is or could be, and further development of the practices and potential research directions we’ve outlined here.

ACKNOWLEDGMENTS

This work was partly funded by Digital Futures, and the Wallenberg AI, Autonomous Systems and Software Program—Humanities and Society (WASP-HS) funded by the Marianne and Marcus Wallenberg Foundation and the Marcus and Amalia Wallenberg Foundation.

REFERENCES

- [1] 2022. RO-MAN22 GenR Workshop - Keynote Speaker. <https://sites.google.com/view/ro-man-genr-workshop/keynote-speaker>.
- [2] 2022. Robots Need the Ability to Navigate Abusive Interactions.
- [3] Jose Abdelnour-Nocera, Torkil Clemmensen, and Masaaki Kurosu. 2013. Reframing HCI through local and indigenous perspectives. *International Journal of Human-Computer Interaction* 29, 4 (2013), 201–204.
- [4] Sara Ahmed. 2017. *Living a feminist life*. Duke University Press. <http://www.jstor.org/stable/j.ctv11g9836>
- [5] Sareeta Amrute. 2016. *Encoding race, encoding class: Indian IT workers in Berlin*. Duke University Press.
- [6] Stephanie Arevalo Arboleda, Max Pascher, Annalies Baumeister, Barbara Klein, and Jens Gerken. 2021. Reflecting upon Participatory Design in Human-Robot Collaboration for People with Motor Disabilities: Challenges and Lessons Learned from Three Multiyear Projects. In *The 14th Pervasive Technologies Related to Assistive Environments Conference*. 147–155.
- [7] Kagonyi Awori, Frank Vetere, and Wally Smith. 2016. Sessions with grandma: fostering indigenous knowledge through video mediated communication. In *Proceedings of the First African Conference on Human Computer Interaction*. 1–11.
- [8] Madeline Balaam, Rob Comber, Ed Jenkins, Selina Sutton, and Andrew Garbett. 2015. FeedFinder: A Location-Mapping Mobile Application for Breastfeeding Women. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) (CHI '15). Association for Computing Machinery, New York, NY, USA, 1709–1718. <https://doi.org/10.1145/2702123.2702328>
- [9] Karen Barad. 2007. *Meeting the universe halfway: Quantum physics and the entanglement of matter and meaning*. duke university Press.
- [10] 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* (CHI '10). Association for Computing Machinery, New York, NY, USA, 1301–1310. <https://doi.org/10.1145/1753326.1753521>
- [11] Ruha Benjamin. 2019. Race after technology: Abolitionist tools for the new jim code. *Social Forces* (2019).
- [12] Tapomayukh Bhattacharjee, Ethan K Gordon, Rosario Scalise, Maria E Cabrera, Anat Caspi, Maya Cakmak, and Siddhartha S Srinivasa. 2020. Is more autonomy always better? exploring preferences of users with mobility impairments in robot-assisted feeding. In *2020 15th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 181–190.
- [13] Alexis E. Block, Hasti Seifi, Otmar Hilliges, Roger Gassert, and Katherine J. Kuchenbecker. 2022. In the Arms of a Robot: Designing Autonomous Hugging Robots with Intra-Hug Gestures. *ACM Transactions on Human-Robot Interaction* (Feb. 2022). <https://doi.org/10.1145/3526110> Just Accepted.
- [14] Alan Borning and Michael Muller. 2012. Next steps for value sensitive design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '12). Association for Computing Machinery, New York, NY, USA, 1125–1134. <https://doi.org/10.1145/2207676.2208560>
- [15] Joy Buolamwini and Timnit Gebru. 2018. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency*. PMLR, 77–91.
- [16] Judith Butler. 2011. *Bodies That Matter: On the Discursive Limits of Sex*. Routledge, London. <https://doi.org/10.4324/9780203828274>
- [17] Wan-Ling Chang and Selma Šabanović. 2015. Interaction Expands Function: Social Shaping of the Therapeutic Robot PARO in a Nursing Home. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI '15)*. ACM, New York, NY, USA, 343–350. <https://doi.org/10.1145/2696454.2696472>
- [18] Sumi Cho, Kimberlé Williams Crenshaw, and Leslie McCall. 2013. Toward a Field of Intersectionality Studies: Theory, Applications, and Praxis. *Signs: Journal of Women in Culture and Society* 38, 4 (June 2013), 785–810. <https://doi.org/10.1086/669608> Publisher: The University of Chicago Press.
- [19] Patricia Hill Collins. 2002. *Black feminist thought: Knowledge, consciousness, and the politics of empowerment*. routledge.
- [20] Julia R Cordero, Thomas R Groechel, and Maja J Matarić. 2022. What and How Are We Reporting in HRI? A Review and Recommendations for Reporting Recruitment, Compensation, and Gender. *arXiv preprint arXiv:2201.09114* (2022).
- [21] Sasha Costanza-Chock. 2018. Design Justice, A.I., and Escape from the Matrix of Domination. *Journal of Design and Science* (July 2018). <https://doi.org/10.21428/96c8d426>
- [22] Kimberlé Crenshaw. 1997. Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics. In *Feminist Legal Theories*. Routledge. Num Pages: 29.
- [23] Kimberlé Williams Crenshaw. 1994. Mapping the Margins: Intersectionality, Identity Politics, and Violence Against Women of Color. In *The Public Nature of Private Violence*. Routledge. Num Pages: 26.
- [24] Simone de Beauvoir. 2014. *The second sex*. Routledge. <https://doi.org/10.4324/9781315840154-23> Pages: 118–123 Publication Title: Classic and Contemporary Readings in Sociology.
- [25] Maartje M. A. de Graaf, Giulia Perugia, Eduard Fosch-Villaronga, Angelica Lim, Frank Broz, Elaine Schaertl Short, and Mark Neerincx. 2022. Inclusive HRI: Equity and Diversity in Design, Application, Methods, and Community. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 1247–1249. <https://doi.org/10.1109/HRI53351.2022.9889455>
- [26] Ewerton de Oliveira, Laura Donadoni, Stefano Boriero, and Andrea Bonarini. 2021. Deceptive Actions to Improve the Attribution of Rationality to Playing Robotic Agents. *International Journal of Social Robotics* 13, 2 (April 2021), 391–405. <https://doi.org/10.1007/s12369-020-00647-8>
- [27] Amelia DeFalco. 2020. Towards a theory of posthuman care: Real humans and caring robots. *Body & Society* 26, 3 (2020), 31–60.
- [28] Catherine D'Ignazio and Lauren F. Klein. 2020. *Data Feminism*. MIT Press.
- [29] Skyla Y. Dudek and James E. Young. 2022. Fluid Sex Robots: Looking to the 2LGBTQIA+ Community to Shape the Future of Sex Robots. In *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction (HRI '22)*. IEEE Press, Sapporo, Hokkaido, Japan, 746–749.
- [30] Saad Elbeleidi, Terran Mott, and Tom Williams. 2022. Practical, Ethical, and Overlooked: Teleoperated Socially Assistive Robots in the Quest for Autonomy. In *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction (HRI '22)*. IEEE Press, Sapporo, Hokkaido, Japan, 577–587.
- [31] Emma Engstrom and Karim Jebari. 2022. AI4People or People4AI? On human adaptation to AI at work. *AI & SOCIETY* (June 2022). <https://doi.org/10.1007/s00146-022-01464-5>
- [32] Friederike Eyssel. 2022. What's Social about Social Robots? A Psychological Perspective. In *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction (HRI '22)*. IEEE Press, Sapporo, Hokkaido, Japan, 2.
- [33] Wendy Faulkner. 2000. Dualisms, hierarchies and gender in engineering. *Social studies of science* 30, 5 (2000), 759–792.
- [34] Heike Felzmann, Eduard Fosch-Villaronga, Christoph Lutz, and Aurelia Tamar-Larriex. 2019. Robots and transparency: The multiple dimensions of transparency in the context of robot technologies. *IEEE Robotics & Automation Magazine* 26, 2 (2019), 71–78.
- [35] Janet Finch, Professor Janet V. Finch, and Jennifer Mason. 1992. *Negotiating Family Responsibilities*. Routledge, London. <https://doi.org/10.4324/9780203393208>
- [36] 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*. ACM, 129–136.
- [37] Eduard Fosch-Villaronga and Adam Poulsen. 2021. Sex Robots in Care: Setting the Stage for a Discussion on the Potential Use of Sexual Robot Technologies for Persons with Disabilities. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (HRI '21 Companion)*. Association for Computing Machinery, New York, NY, USA, 1–9. <https://doi.org/10.1145/3434074.3446907>
- [38] E. Fosch-Villaronga, A. Poulsen, R. A. Søraa, and B. H. M. Custers. 2021. A little bird told me your gender: Gender inferences in social media. *Information Processing & Management* 58, 3 (May 2021), 102541. <https://doi.org/10.1016/j.ipm.2021.102541>
- [39] Mary Ellen Foster, Andre Gaschler, and Manuel Giuliani. 2017. Automatically Classifying User Engagement for Dynamic Multi-party Human-Robot Interaction. *International Journal of Social Robotics* (July 2017), 1–16. <https://doi.org/10.1007/s12369-017-0414-y>
- [40] Sarah Gillet, Ronald Cumbal, André Pereira, José Lopes, Olov Engwall, and Iolanda Leite. 2021. Robot Gaze Can Mediate Participation Imbalance in Groups with Different Skill Levels. In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (HRI '21)*. Association for Computing Machinery, New York, NY, USA, 303–311. <https://doi.org/10.1145/3434073.3444670>
- [41] Sarah Gillet and Iolanda Leite. 2020. A Robot Mediated Music Mixing Activity for Promoting Collaboration among Children. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI '20)*. Association for Computing Machinery, Cambridge, United Kingdom, 212–214. <https://doi.org/10.1145/3371382.3378307>
- [42] Hatice Gunes, Oya Celiktutan, Evangelos Sariyanidi, and Efstratios Skordos. 2015. Real-time Prediction of User Personality for Social Human-Robot Interactions: Lessons Learned from Public Demonstrations. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) Workshops*.
- [43] Judith Halberstam. 2017. F2M: The making of female masculinity. In *Feminist theory and the body*. Routledge, 125–133.
- [44] Donna J Haraway. 1997. *Feminism and technoscience*. New York: Routledge 39, 1 (1997), 3.
- [45] Sandra G Harding. 1986. *The science question in feminism*. Cornell University Press.
- [46] Mar Hicks. 2017. *Programmed inequality: How Britain discarded women technologists and lost its edge in computing*. MIT Press.
- [47] Marie Hicks. 2019. Hacking the Cis-tem. *IEEE Annals of the History of Computing* 41, 1 (2019), 20–33.
- [48] James Hodge, Sarah Foley, Rens Brankaert, Gail Kenning, Amanda Lazar, Jennifer Boger, and Kellie Morrissey. 2020. Relational, Flexible, Everyday: Learning from Ethics in Dementia Research. In *Proceedings of the 2020 CHI Conference on*

- Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–16. <https://doi.org/10.1145/3313831.3376627>
- [49] Sarah Homewood, Laurens Boer, and Anna Vallgård. 2020. Designers in White Coats: Deploying Ovum, a Fertility Tracking Device. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3313831.3376528>
- [50] Kristina Hook. 2018. *Designing with the Body: Somaesthetic Interaction Design*. MIT Press.
- [51] bell hooks. 2000. *Feminist Theory: From Margin to Center*. Pluto Press.
- [52] Quirien R. M. Hover, Ella Velner, Thomas Beelen, Mieke Boon, and Khiet P. Truong. 2021. Uncanny, Sexy, and Threatening Robots: The Online Community's Attitude to and Perceptions of Robots Varying in Humanlikeness and Gender. In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (HRI '21)*. Association for Computing Machinery, New York, NY, USA, 119–128. <https://doi.org/10.1145/3434073.3444661>
- [53] Wei Huang. 2015. When HCI Meets HRI: the intersection and distinction. *Virginia Polytechnic Institute and State University*, January (2015).
- [54] Emily Ip, Wynnie Wing Yi Chung, Sunmin Lee, and Thecla Schiphorst. 2014. The wearable self: Braiding a feminist critique within a somaesthetics framework for design. In *International Conference of Design, User Experience, and Usability*. Springer, 285–296.
- [55] Lauren Michele Jackson. 2018. Shudu Gram Is a White Man's Digital Projection of Real-Life Black Womanhood. *The New Yorker* (May 2018). <https://www.newyorker.com/culture/culture-desk/shudu-gram-is-a-white-mans-digital-projection-of-real-life-black-womanhood>
- [56] Sooyeon Jeong, Deirdre E. Logan, Matthew S. Goodwin, Suzanne Graca, Brianna O'Connell, Honey Goodenough, Laurel Anderson, Nicole Stenquist, Katie Fitzpatrick, Miriam Zisook, Luke Plummer, Cynthia Breazeal, and Peter Weinstock. 2015. A Social Robot to Mitigate Stress, Anxiety, and Pain in Hospital Pediatric Care. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts*. ACM, Portland Oregon USA, 103–104. <https://doi.org/10.1145/2701973.2702028>
- [57] Candace L. Kemp, Mary M. Ball, and Molly M. Perkins. 2013. Convoys of care: Theorizing intersections of formal and informal care. *Journal of Aging Studies* 27, 1 (Jan. 2013), 15–29. <https://doi.org/10.1016/j.jaging.2012.10.002>
- [58] Alex Ketchum. 2021. Intro to Feminist and Social Justice Studies - Episode 17: Technology (2021). <https://introtofeministstudies.blogspot.com/2020/07/episode-17-technology.html>
- [59] Os Keyes. 2018. The misgendering machines: Trans/HCI implications of automatic gender recognition. *Proceedings of the ACM on human-computer interaction* 2, CSCW (2018), 1–22.
- [60] Boyoung Kim, Ruchen Wen, Qin Zhu, Tom Williams, and Elizabeth Phillips. 2021. Robots as Moral Advisors: The Effects of Deontological, Virtue, and Confucian Role Ethics on Encouraging Honest Behavior. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (HRI '21 Companion)*. Association for Computing Machinery, New York, NY, USA, 10–18. <https://doi.org/10.1145/3434074.3446908>
- [61] Debora de Castro Leal, Angelika Strohmayer, and Max Krüger. 2021. On activism and academia: Reflecting together and sharing experiences among critical friends. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–18.
- [62] Séverin Lemaignan, Fernando Garcia, Alexis Jacq, and Pierre Dillenbourg. 2016. From Real-time Attention Assessment to "With-me-ness" in Human-Robot Interaction. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction (HRI '16)*. IEEE Press, Piscataway, NJ, USA, 157–164. <http://dl.acm.org/citation.cfm?id=2906831.2906860>
- [63] Risa Maeda, Dražen Brščić, and Takayuki Kanda. 2021. Influencing Moral Behavior Through Mere Observation of Robot Work: Video-Based Survey on Littering Behavior. In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (HRI '21)*. Association for Computing Machinery, New York, NY, USA, 83–91. <https://doi.org/10.1145/3434073.3444680>
- [64] Emily Martin. 1991. The egg and the sperm: How science has constructed a romance based on stereotypical male-female roles. *Signs: Journal of Women in Culture and Society* 16, 3 (1991), 485–501.
- [65] Derek McColl and Goldie Nejat. 2014. Determining the affective body language of older adults during socially assistive HRI. In *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on*. IEEE, 2633–2638.
- [66] Rajiv Mehta and Dawn Nafus. 2016. *Atlas of Caregiving Pilot Study Report*. Technical Report.
- [67] Gaspar Isaac Melsion, Ilaria Torre, Eva Vidal, and Iolanda Leite. 2021. Using Explainability to Help Children Understand Gender Bias in AI. In *Interaction Design and Children*. 87–99.
- [68] Reijo Miettinen. 1999. The riddle of things: Activity theory and actor-network theory as approaches to studying innovations. *Mind, Culture, and Activity* 6, 3 (1999), 170–195.
- [69] Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. Model Cards for Model Reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '19)*. Association for Computing Machinery, New York, NY, USA, 220–229. <https://doi.org/10.1145/3287560.3287596>
- [70] Hannah Moore, Lucy Osborne, Rose de Larrabeiti, Rudi Zygadlo, Phil Maynard, and Elizabeth Cassin. 2022. Abuse in the Modelling Industry – Podcast. *the Guardian* (May 2022). <https://www.theguardian.com/news/audio/2022/may/31/breaking-silence-brunel-abuse-modelling-industry-podcast>
- [71] Fatemeh Moradi, Linnea Öhlund, Hanna Nordin, and Mikael Wiberg. 2020. Designing a Digital Archive for Indigenous People: Understanding the Double Sensitivity of Design. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*. 1–11.
- [72] Alistair Morrison, Donald McMillan, and Matthew Chalmers. 2014. Improving Consent in Large Scale Mobile HCI through Personalised Representations of Data. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational*. ACM, 471–480.
- [73] Kayako Nakagawa, Masahiro Shiomi, Kazuhiko Shinozawa, Reo Matsumura, Hiroshi Ishiguro, and Norihiro Hagita. 2011. Effect of robot's active touch on people's motivation. In *Human-Robot Interaction (HRI), 2011 6th ACM/IEEE International Conference on*. IEEE, 465–472.
- [74] Jennifer C. Nash. 2018. *Black Feminism Reimagined: After Intersectionality*. Duke University Press.
- [75] Sarah Ng, Shaowen Bardzell, and Jeffrey Bardzell. 2020. The Menstruating Entrepreneur Kickstarting a New Politics of Women's Health. *ACM Transactions on Computer-Human Interaction (TOCHI)* 27, 4 (2020), 1–25.
- [76] Ihudiya Finda Ogbonnaya-Ogburu, Angela D.R. Smith, Alexandra To, and Kentaro Toyama. 2020. Critical Race Theory for HCI. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–16. <https://doi.org/10.1145/3313831.3376392>
- [77] Aude Oliva, Ben Armstrong, Emily Goldman, David Kaiser, and Milo Philips-Brown. 2022. MIT Ethical Computing Platform | Tools & Resources to Help Make Technology More Responsible. <https://comp-wp-spare-2.mit.edu/>.
- [78] Nelly Oudshoorn. 2003. *Beyond the natural body: An archaeology of sex hormones*. Routledge.
- [79] Sharon Oviatt. 2021. Technology as Infrastructure for Dehumanization: Three Hundred Million People with the Same Face. In *Proceedings of the 2021 International Conference on Multimodal Interaction (Montréal, QC, Canada) (ICMI '21)*. Association for Computing Machinery, New York, NY, USA, 278–287. <https://doi.org/10.1145/3462244.3482855>
- [80] Jennifer A. Parks. 2003. *No Place Like Home?: Feminist Ethics and Home Health Care*. Indiana University Press, Bloomington. <https://muse.jhu.edu/pub/3/monograph/book/9154>
- [81] Giulia Perugia, Stefano Guidi, Margherita Bicchì, and Oronzo Parlangei. 2022. The Shape of Our Bias: Perceived Age and Gender in the Humanoid Robots of the ABOT Database. In *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction (HRI '22)*. IEEE Press, Sapporo, Hokkaido, Japan, 110–119.
- [82] Aaron Powers. 2008. FEATURE What robotics can learn from HCI. *Interactions* 15, 2 (2008), 67–69.
- [83] Jennifer Rhee. 2018. All Too Human: Labor and Dehumanization in the Robotic Imaginary.
- [84] Laurel Riek and Don Howard. 2014. A code of ethics for the human-robot interaction profession. *Proceedings of we robot* (2014).
- [85] CM Roberts. 2007. *Messengers of Sex: Hormones, biomedicine and feminism*. Cambridge University Press.
- [86] Jennifer Robertson. 2017. *Robo sapiens japonicus*. University of California Press.
- [87] Astrid M. Rosenthal-von der Pütten, Nicole C. Krämer, Laura Hoffmann, Sabrina Sobieraj, and Sabrina C. Eimler. 2013. An Experimental Study on Emotional Reactions Towards a Robot. *International Journal of Social Robotics* 5, 1 (Jan. 2013), 17–34. <https://doi.org/10.1007/s12369-012-0173-8>
- [88] Nina J. Rothstein, Dalton H. Connolly, Ewart J. de Visser, and Elizabeth Phillips. 2021. Perceptions of Infidelity with Sex Robots. In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (HRI '21)*. Association for Computing Machinery, New York, NY, USA, 129–139. <https://doi.org/10.1145/3434073.3444653>
- [89] Selma Sabanovic. 2010. Robots in Society, Society in Robots. *International Journal of Social Robotics* 2, 4 (Dec. 2010), 439–450. <https://doi.org/10.1007/s12369-010-0066-7>
- [90] Morgan Klaus Scheuerman, Katta Spiel, Oliver L. Haimson, Foad Hamidi, and Stacy M. Branham. 2020. HCI guidelines for gender equity and inclusivity. *UMBC Faculty Collection* (2020).
- [91] Matthias Scheutz and Thomas Arnold. 2016. Are We Ready for Sex Robots?. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction (HRI '16)*. IEEE Press, Christchurch, New Zealand, 351–358.
- [92] Ari Schlesinger, Kenton P. O'Hara, and Alex S. Taylor. 2018. Let's talk about race: Identity, chatbots, and AI. In *Proceedings of the 2018 chi conference on human factors in computing systems*. 1–14.

- [93] J. Shen, O. Rudovic, S. Cheng, and M. Pantic. 2015. Sentiment apprehension in human-robot interaction with NAO. In *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*. 867–872. <https://doi.org/10.1109/ACII.2015.7344676>
- [94] Margrit Shildrick. 2015. *Leaky bodies and boundaries: Feminism, postmodernism and (bio) ethics*. Routledge.
- [95] Masahiro Shiomi, Aya Nakata, Masayuki Kanbara, and Norihiro Hagita. 2017. A hug from a robot encourages prosocial behavior. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. 418–423. <https://doi.org/10.1109/ROMAN.2017.8172336> ISSN: 1944-9437.
- [96] Masahiro Shiomi, Aya Nakata, Masayuki Kanbara, and Norihiro Hagita. 2017. A Robot that Encourages Self-disclosure by Hug. In *Social Robotics (Lecture Notes in Computer Science)*, Abderrahmane Kheddar, Eiichi Yoshida, Shuzhi Sam Ge, Kenji Suzuki, John-John Cabibihan, Friederike Eyssel, and Hongsheng He (Eds.). Springer International Publishing, Cham, 324–333. https://doi.org/10.1007/978-3-319-70022-9_32
- [97] Marie Louise Juul Søndergaard, Ozgun Kilic Afsar, Marianela Ciolfi Felice, Nadia Campo Woytuk, and Madeline Balaam. 2020. Designing with Intimate Materials and Movements: Making "Menarche Bits". In *Proceedings of the 2020 ACM Designing Interactive Systems Conference (Eindhoven, Netherlands) (DIS '20)*. Association for Computing Machinery, New York, NY, USA, 587–600. <https://doi.org/10.1145/3357236.3395592>
- [98] Roger Andre Søraa. 2017. Mechanical genders: how do humans gender robots? *Gender, Technology and Development* 21, 1-2 (2017), 99–115.
- [99] Robert Sparrow. 2020. Robotics has a race problem. *Science, Technology, & Human Values* 45, 3 (2020), 538–560.
- [100] Katta Spiel. 2021. "Why are they all obsessed with Gender?" — (Non)binary Navigations through Technological Infrastructures. In *Designing Interactive Systems Conference 2021*. ACM, Virtual Event USA, 478–494. <https://doi.org/10.1145/3461778.3462033>
- [101] Katta Spiel, Emeline Brulé, Christopher Frauenberger, Gilles Bailly, and Geraldine Fitzpatrick. 2018. Micro-ethics for participatory design with marginalised children. In *Proceedings of the 15th Participatory Design Conference: Full Papers - Volume 1 (PDC '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. <https://doi.org/10.1145/3210586.3210603>
- [102] Anna Ståhl, Vasiliki Tsaknaki, and Madeline Balaam. 2021. Validity and Rigour in Soma Design-Sketching with the Soma. *ACM Trans. Comput.-Hum. Interact.* 28, 6, Article 38 (dec 2021), 36 pages. <https://doi.org/10.1145/3470132>
- [103] Megan Strait, Florian Lier, Jasmin Bernotat, Sven Wachsmuth, Friederike Eyssel, Robert Goldstone, and Selma Šabanović. 2020. A three-site reproduction of the joint simon effect with the NAO robot. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*. 103–111.
- [104] Yolande Strengers and Jenny Kennedy. 2021. *The smart wife: Why Siri, Alexa, and other smart home devices need a feminist reboot*. MIT Press.
- [105] Angelika Strohmayr, Janis Lena Meissner, Alexander Wilson, Sarah Charlton, and Laura McIntyre. 2020. "We Come Together as One...and Hope for Solidarity to Live on": On Designing Technologies for Activism and the Commemoration of Lost Lives. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference (Eindhoven, Netherlands) (DIS '20)*. Association for Computing Machinery, New York, NY, USA, 87–100. <https://doi.org/10.1145/3357236.3395452>
- [106] Lucy Suchman and Lucy A Suchman. 2007. *Human-machine reconfigurations: Plans and situated actions*. Cambridge university press.
- [107] Cara Tannenbaum, Robert P. Ellis, Friederike Eyssel, James Zou, and Londa Schiebinger. 2019. Sex and gender analysis improves science and engineering. *Nature* 575, 7781 (Nov. 2019), 137–146. <https://doi.org/10.1038/s41586-019-1657-6> Number: 7781 Publisher: Nature Publishing Group.
- [108] Laetitia Tanqueray, Tobiaz Paulsson, Mengyu Zhong, Stefan Larsson, and Ginevra Castellano. 2022. Gender Fairness in Social Robotics: Exploring a Future Care of Peripartum Depression. In *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction (HRI '22)*. IEEE Press, Sapporo, Hokkaido, Japan, 598–607.
- [109] Hamish Tennent, Solace Shen, and Malte Jung. 2019. Micbot: A Peripheral Robotic Object to Shape Conversational Dynamics and Team Performance. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 133–142. <https://doi.org/10.1109/HRI.2019.8673013> ISSN: 2167-2148.
- [110] Sharon Traweek. 2009. *Beamtimes and lifetimes*. Harvard University Press.
- [111] Pat Treusch, Arne Berger, and Daniela K Rosner. 2020. Useful uselessness? Teaching robots to knit with humans. In *Proceedings of the 2020 ACM designing interactive systems conference*. 193–203.
- [112] Ralph Vacca. 2021. Value-Sensitive Design for Feminist Technology: Designing for "Feminist Fatherhood" with Nonwhite Latinx and White Fathers. *Global Perspectives* 2, 1 (Feb. 2021). <https://doi.org/10.1525/gp.2021.24917> Publisher: University of California Press.
- [113] Stephanie Valencia, Michal Luria, Amy Pavel, Jeffrey P. Bigham, and Henny Admoni. 2021. Co-designing Socially Assistive Sidekicks for Motion-based AAC. In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (HRI '21)*. Association for Computing Machinery, New York, NY, USA, 24–33. <https://doi.org/10.1145/3434073.3444646>
- [114] Maja Van der Velden. 2013. Decentering design: Wikipedia and indigenous knowledge. *International Journal of Human-Computer Interaction* 29, 4 (2013), 308–316.
- [115] Judy Wajcman. 2004. *Technofeminism*. Cambridge: Polity. (2004).
- [116] Mark West, Rebecca Kraut, and Han Ei Chew. 2019. *I'd Blush If I Could: Closing Gender Divides in Digital Skills through Education*. Technical Report.
- [117] Katie Winkle, Erik Lagerstedt, Ilaria Torre, and Anna Offenwanger. 2022. 15 Years of (Who)man Robot Interaction: Reviewing the H in Human-Robot Interaction. *ACM Transactions on Human-Robot Interaction* (Nov. 2022). <https://doi.org/10.1145/3571718>
- [118] Katie Winkle, Gaspar Isaac Melsión, Donald McMillan, and Iolanda Leite. 2021. Boosting Robot Credibility and Challenging Gender Norms in Responding to Abusive Behaviour: A Case for Feminist Robots. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*. 29–37.
- [119] Robert H Wortham, Andreas Theodorou, and Joanna J Bryson. 2016. What does the robot think? Transparency as a fundamental design requirement for intelligent systems. In *Ijcai-2016 ethics for artificial intelligence workshop*.
- [120] Yuchen Yan and Yunyi Jia. 2022. A Review on Human Comfort Factors, Measurements, and Improvements in Human–Robot Collaboration. *Sensors* 22, 19 (2022), 7431.
- [121] Josie Young. 2020. Feminist Design Tool Defensible Decision Making for Interaction Design and AI.
- [122] Qin Zhu, Tom Williams, Blake Jackson, and Ruchen Wen. 2020. Blame-laden moral rebukes and the morally competent robot: A Confucian ethical perspective. *Science and Engineering Ethics* 26, 5 (2020), 2511–2526.
- [123] Megan Zimmerman, Shelly Bagchi, Jeremy Marvel, and Vinh Nguyen. 2022. An Analysis of Metrics and Methods in Research from Human-Robot Interaction Conferences, 2015-2021. In *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction (HRI '22)*. IEEE Press, Sapporo, Hokkaido, Japan, 644–648.
- [124] Douglas Zytco and Wing-Yue Geoffrey Louie. 2022. Ethical Considerations When Constructing Participatory Design Protocols for Social Robots. *arXiv preprint arXiv:2203.10398* (2022).