What Norms Are Social Robots Reflecting?
A Socio-Legal Exploration on HRI Developers

Laetitia TANQUERAY and Stefan LARSSON
Department of Technology and Society, Lund University, Sweden

Abstract. By relying on theory from sociology of law and data feminism, this study showcases the norms guiding development in human-robot interaction. This qualitative study consists of an ethnography of the HRI conference 2021 and expert interviews which were merged and analysed using an ethnographic content analysis method on NVivo. The socio-legal data feminist lens enables to pinpoint the lack of clear legal involvement, the reliance on the HRI community to develop, and the normative impact this has on the overall development of social robots. This study not only aims to showcase the vital role of HRI developers, but also the need for more critical scholars in this area.

Keywords. sociology of law, data feminism, human-robot interaction (HRI) developers, social robots

1. Introduction: Norms and HRI

Recent development in artificial intelligence (AI) has sparked lively debates on how to make AI more ethical, more trustworthy, and less biased (for example [1, 2]). A range of principled guidelines suggest bringing more disciplines into the design of AI; although this may also lead to some type of confusion where the engineers creating AI tools are conflicted about which guidelines to follow [3], or what calls for transparency or fairness actually means in detail or for specific contexts [4, 5, 6]. This paper places these ideas within the Human-Robot Interaction (HRI) field, by empirically focusing on HRI developers, which is primarily made up of engineers. Accordingly, this paper underscores the who question rather than the how question on the development of new digital technologies. This is achieved by conducting an ethnography of the HRI conference and interviews with social roboticists, through a socio-legal feminist qualitative lens.

1.1. Purpose

The purpose of this study is to understand what norms are reflected when designing social robots, by focusing on the HRI community from a socio-legal feminist perspective. The
outcome is intended to showcase the developing steering norms HRI, and the need for more critical studies to help inform HRI design.

By focusing on the who questions, we investigate who are behind making social robots and what the potential normative impacts they have on social robots. Hence the title “what norms are social robots reflecting”. It is directly influenced by a long tradition in the discipline of Sociology of Law (SoL), which explores the concept of norms, particularly informal structures of behavioural expectations [7, 8], and the power these norms may have in a certain context, to pinpoint the underlying social structures at play and their consequences. To accommodate for this, we explore the HRI community through combining SoL and data feminism. In particular, we draw from SoL scholar Larsson [9]’s concept of mirroring norms, which questions the social structures within the data and the difficult normative choices that developers make inadvertently. Data feminism, in addition, offers a critical lens in data science drawn up by D’Ignazio and Klein’s [10]; two concepts of which are of interest for this study: privilege hazard and stranger in the dataset, both of which account for the vital role developers play.

1.2. What Norms Do Robots Reflect?

Applying mirroring norms, privilege hazard, and stranger in the dataset is of interest here due to the very nature of social robots. Social robots are physical entities with personalities and characters, who can perceive and express emotions as well as communicate through the use of natural cues, such as gaze and gestures [11, p.145] [12, p.443], and act according to expected behaviour norms within a certain context [13, 14] – all of which may use artificial intelligence (AI) to adapt somewhat autonomously to the user. From this definition, and by drawing from insights in critical design studies [10, 15], it is clear that HRI developers have to make important normative decisions – unwittingly or not. Thus, by situating itself at the development of social robots, this paper explores the HRI community by observing an acclaimed social robotics conference, as well as interviewing social roboticists.

We divide norms into two branches of significance here: legal norms and social norms, which guides individual and collective behaviours [17]. This may be seen as a spectrum or overlap, as social norms can be formalised into legal norms, or work as a supportive informal structure. They may also be in conflict and exist in parallel [18]. Legal norms can be viewed as formalised ”ought” statements [17], and for the sake of this analysis, we distinguish between two layers of formal legal norms which emanate from: (1) enforceable laws, such as the GDPR regarding data protection in the European context [19] and (2) non-enforceable guides, such as recommendations, guidelines and standards, usually created by industry representatives of the interest groups or governing bodies, for instance the IEEE Guidelines or the Ethics Guidelines on AI [1, 2]. At the core of this paper is the focus on the relationship between social and legal norm, as this is what ontologically shapes mirroring norms; whereby there are informal guiding social norms, which derive from outside of the legal realm, yet will guide and influence development in HRI.

2 The empirical data and findings are based on the primary author’s research for her master’s thesis (Available at: url=http://lup.lub.lu.se/student-papers/record/9065319 [16]).
2. Relevant Work Informing the Theoretical Framework

In this section we draw from studies on social structures, based on a theoretical foundation with quite a long prehistory in the social sciences, to combine this with more recent insights in data-dependent design and critical perspectives on AI. This combination serves as the theoretical framework for this study.

2.1. Social Structures and Norms

Social structures embedded in new technology, including social robots and AI, have been critically assessed from various perspectives, including: race [20], gender [15], queer community [21], as well as the binary oversimplification of individuals [22]. This work often points to the imbalances within society outside of technology, and yet how technology may amplify or reproduce them [9].

The role developers play generally is discussed by critical scholars to demonstrate how it is possible to account for the diversity of people [23, 24, 10]. Although this work is invaluable, engineers creating new technologies have spoken out about their role evolving from technical development, to now also having to account for broader social context [3, p.7]. The latter issue, for example, has been brought up by engineers in a hackathon to bring their own views to AI guideline developments [3]. They discuss how these guidelines do not account for engineers’ expertise which is that of innovators, as those guidelines also add a burden on them to take into account societal implications – despite engineers not being trained to do so. This was also echoed by Fosch-Villaronga et al [12] when gathering expert opinions on social robots; they found that guidelines did not offer certainty, especially not legal certainty to developers [12, p.447].

Specifically developers of social robots have been scrutinised by Richardson [25] through observing developers creating social robots at MIT; Strengers and Kennedy [26] mapping out the discourse behind social robots and its significance; and Šabanović [27] a social roboticist showcasing how developers in Japan have to follow conservative values so that the robot will be accepted within everyday society. Tanqueray et al [28] represent a collaboration between SoL and HRI scholars. This collaboration enabled to outline norms within the context of peripartum depression, and how this can be incorporated into the design of social robots to screen for peripartum depression [28].

2.2. Interdisciplinary Framework: Sociology of Law and Data Feminism

Theories in SoL regarding new digital technologies are still nascent. Thus this paper relies on two different disciplines: SoL, by integrating Larsson’s concept of mirroring norms [9], and data feminism, through D’Ignazio and Klein’s concepts of privilege hazard and stranger in the dataset [10].

The social structures and norms emanating from those social structures are key to SoL as a discipline. To the founding fathers of SoL, the normative structures in society not only include legal norms but also informal “laws” or norms – the living law (Ehrlich) or intuitive law (Petrazyki). Accordingly, researchers in SoL often study aspects of social norms, and their relation to legal norms, within society. Larsson [9] and Hydén [8, 7], both SoL scholars, lean on these informal laws to demonstrate the pivotal role those norms play within digital and technological developments in data and to developers. In-
formal "laws" make up norms which become part of interactions between people and institutions, which are in turn reflected in new adaptive technologies through collected data which makes up datasets. Mirroring norms thus refers to the behind the scenes of adaptive technologies. It recognises that the social structures within datasets risk echoing not only "the balanced sides of humanity" but also the "biased, skewed and discriminatory" sides of society [9, p.575]. Thus, it is for developers to take normative positions even if they do not wish to do so [9, p.591] as the data inherently mirrors normatively (including problematic) social norms.

In parallel, D’Ignazio and Klein highlight problematic implications of developers’ blind spots within data science from an intersectional feminist standpoint. Their work illustrates how to challenge current data science practices, of which are privilege hazard and stranger in the dataset, to bring awareness to issues around collecting and classifying data. Privilege hazard as a concept acknowledges that developers usually occupy the most privileged position in society, through education and credentials, however this creates an obstacle for them to “recognise instances of oppression in the world” when shaping and scaling up data [10, p.29]. Stranger in the dataset follows on from this. It identifies developers as having to make sense of the collected data despite not being part of the community; this leads to developers having to make normative decisions which might not recognise the tainted history or the social structures they reproduce [10, p.131].

Through these three concepts, we demonstrate the intricacies of the role of developers generally; by pointing to what HRI developers might mirror and the consequences of their explorations and involvement when automating processes in social robots.

3. Methodology

The research design has as its object the developers of social robots, which are recognised as experts in their field. In line with Döringer, “expert” in this context is meant for people working on developing social robots, however none of them have much (if any) knowledge on or consideration for SoL [29, p.265]. In line with this, the primary author collected empirical data from experts in social robotics, first through a week-long ethnography of the online HRI conference and second by conducting expert interviews. The conference is deemed prestigious in the HRI community as it showcases the latest cutting-edge research in HRI [30]. Accordingly, note-keeping and reflections of the conference from the first author made up part of the data analysed [31]. After this, three semi-structured in-depth interviews following a strict interview guide were conducted by the first author (between 60 to 115 minutes through video-call) with HRI experts within the industry (3 Participants; Male: 1, Female: 2). They are referred to as P1, P2, P3 in the findings.

All the data was merged together in NVivo and analysed qualitatively using ethnographic content analysis (ECA). This method stems from media and communication studies and allows for the researcher to be central to the coding stage (i.e. the role she played when collecting the data and the iterative coding process) [32]. This was key since the authors are not educated in HRI and viewing the community from a socio-legal lens. In order to ensure replicability, a code book was created and maintained throughout the coding stage. Most of the codes were predetermined, but were re-adapted through constant reiteration. The code book described the name of the nodes and children nodes as well as a clear delimitation of when to use the code on the data and when not to.
4. Findings

The findings are divided into three sections: section (1) shows the perception of law, divided into (a) how legal norms are understood in the HRI community and (b) the need for more legal competence in HRI; section (2) demonstrates the appreciation of and reliance on the HRI community; and, section (3) showcases the HRI developers’ view of their own role as (a) ”enablers” or (b) ”challengers” of the status quo through robots.

4.1. Where is the Law?

4.1.1. The HRI Community’s Understanding of Law

The HRI community is made up of mostly engineers who can programme robots in a way that will innovately attempt to create a solution to a problem through the use of social robots. Thus, the HRI conference offers a space to demonstrate the possible solutions to specific or general problems to other developers. Nevertheless, the HRI community is also open to understanding how they should inform their design. One of which could emanate from types of enforceable laws and non-enforceable guides.

In line with this, when asked the question in the interviews about what laws or guides developers might follow to inform their design, all responded that they did not know of any except the GDPR. P1 mentioned that the GDPR hindered her research, explaining that it made her research more complicated to collect data to understand humans beyond her own colleagues and herself. P3 joked about the law, saying that because of it there was a random button on the robot, but he did not know why. This was directly reflected within the HRI conference as well. One presentation focused on a specific British Standard to create an ethical robot design (British Standard BS8611). The team of roboticists demonstrated how this Standard, which focused on how to design such a robot, was misaligned with roboticists’ practices and also the users’ view on what an ethical robot might be.

4.1.2. The Lack of Formal Legal Involvement

When interviewing P2, she was vocal about the lack of involvement of policy makers:

Everyone’s like ‘can we democratise AI?’ It’s actually funny because we need to have more lawyers in this discussion. It’s always a bunch of roboticists, engineers, computer scientists and we’re just discussing this among us. Sometimes I think ‘where is the rest of the world?” (P2).

Indeed, the HRI community involves researchers beyond engineering: from the social sciences, it is mostly from psychology and a small proportion of other disciplines such as law and philosophy, to conceptualise HRI issues in their terms. In one of the presentations held by a lawyer for example, they questioned the possible use of sex robots for people with disabilities. The speaker presented various forms of disabilities, which results in different needs for sex robots. The conclusion of this presentation was to make the HRI community aware of sex robots in this sensitive area, which is also overlooked legally, whilst also informing developers on how to ensure mindful designs for the users.

P1 also voiced her worry around her role as a developer when making innovations that have many implications, such as bringing robots into homes or making robots good companions to humans. She compared her role to that of a medical doctor; except medical doctors have strict regulations to follow which also guide them, unlike roboticists.
4.2. **The Mutual Appreciation in HRI**

Roboticists heavily rely on one another’s expertise to be able to bring robots “to life”; for example, one roboticist will concentrate on speech, whilst the other on movement, and both have to be in sync. This constant reliance and appreciation between them is very clear throughout the HRI conference, which, from the interviews and observations, stems from the mutual desire to bring social robots into people’s everyday life. In one instance, one of the speakers shared their code (on an open-source website). The reason was that people who did not feel comfortable coding could see ‘how easy’ it was and how minimal training data was required for certain applications. Indicatively, the interviewees in this study would point to another team that would know more about a certain aspect of the robot to help answer questions the first author had asked them. This was also visible in the way they develop. P1 explained to me that when she is trying to find a solution, she first turns to the HRI conference to inspire from their work. Accordingly, there was a need to rely as well as be amicable, and not compete against one another. This creates a safe space for roboticists to be honest about their failures during the process, although not many openly spoke about those at the conference. This was openly brought up in one presentation at the conference which called out this mutual reliance, and the need to be critical of it before trusting and using the same solution.

4.3. **The Developers’ View of Their Own Role**

4.3.1. **Developers as Enablers**

The interviewees were all aware that the way they develop is a reflection of how they, as developers, view of society, “and there’s no way to get out of that (...) you [the developer] don’t understand every situation” (P3). Moreover, when discussing with interviewees how they choose to develop, one explained how she would spend time with the robot to see its limitation. From there, she would try and find a solution, then try it out on her colleagues as part of her data collection- this practice increased due to the pandemic as she couldn’t get her data from the “real world” (P1). This was implicitly echoed at the conference where presenters allured to the difficulty around data collection due to the pandemic. Nevertheless, most developers in the ethnography and all interviewees viewed their role mostly as enablers of technology in a certain field. Thus, in order to develop in a field that they do not know much about, developers will turn to who they think will be the right person or institution. This showcases how developers see themselves as enablers between the user and the institution that would benefit from it. A good illustration of this was bringing socially assistive robots into hospitals for children. The roboticists asked the medical professionals, first on what could be needed, and second on the prototype the roboticists made based on their comments; as well as asking children their thoughts on the robot. As one interviewee explained, and sums up the example ”the stuff that I would deal with is how does the robot best get the information that the medical researcher wants” (P3).

4.3.2. **Developers as Challengers of the Status Quo**

On the other hand, some developers were questioning their own role, and how they can challenge the status quo. One presenter used the social robot as a mean to challenge the
status quo around gender norms. The robot was programmed to answer back to abuse, either through argumentative or aggressive responses, instead of relying on a standard reply along the lines of "I won’t respond to that". Other developers wanted to ensure children were still able to play with one another during the pandemic through social robots. In this instance, each child in different households could get a robot and play with one another, where the actual cost of the robot was also considered as well as security and privacy issues. Another presentation, focusing on sex robots for people with disabilities, specifically mentioned that they wanted to first present to the community of developers to understand the theory and feasibility, before going to the targeted users and institutions. All these presentations were greeted with great enthusiasm from the HRI attendees, and participants wanted to know how they could help further the research, or reciprocate it, or assist in any way.

5. Applying the Theoretical Framework to the Findings

Applying the theoretical framework presented above to the findings, we discuss (1) the lack of explicit legal norms informing HRI developers, (2) what the impact this has on the development process, and (3) what (unintended) implications this may have on the HRI community.

5.1. The Non-Reflected Formal Legal Norms

Although there are various formal types of laws or normative statements, such as the GDPR and guidelines on AI, the HRI community echoes the literature from understanding engineers and social robotic experts: they feel they have to rely on their own judgement instead of guidelines, as they do not know which one to follow [3, 12]. However, the findings go further by showing that the formal law and non-enforceable law when applicable might be misaligned with developers’ practices, such as the British Standard example. The findings also show that the developers do not necessarily understand the law’s purpose, such as the random button on the robot and the GDPR when undertaking new research. Thus, ideally, formal legal norms would steer developers practices on how to reflect society in a given context [9]. Instead, those formal legal norms are either not there, or there are too many, or they mismatch with actual practices.

5.2. Warning: Mirroring Norms, Privileges, and Strangers

The above shows the lack of formal legal norms steering the development, which results in development being guided through informal but necessary guiding norms. These are therefore enforced by social means, and rely on the judgement of the HRI developers. This has two consequences, (1) an appreciation for each other’s roles, as demonstrated by giving out codes and (2) reliance on one another’s solutions, or using one another to understand how the others understand the issue. This reliance is beginning to be critiqued by the HRI community itself with "challengers". Nevertheless, this appreciation and reliance should not be overlooked, as they make up informal norms. The mirroring norms, privilege hazard and, stranger in the dataset draws on scholars which warn on the important normative stance developers take by default, likely unwittingly, yet they are also by default in a privileged position whereby they do not represent all of society [10, 9].
This privileged position and mirroring each other’s practices, can result in overlooking societal issues and reproducing norms within society that may not be wanted—as studied by a growing number of critical scholars [22, 20, 15, 24, 10, 9, 21, 27].

5.3. The Normative Power Developers Hold

The risks of being strangers in the dataset and mirroring norms, not only focuses on the developers, but also who the developers might involve. This has a significant normative impact on the developing stage of the robot. Indeed, from the findings, developers choose who to collaborate with, and usually this results in collaborating with the powerful institutions—as seen when they collaborated with medical doctors, or testing on each other. Although the HRI community is open to submissions from various backgrounds, as showcased with the lawyer; there was no explicit mention of critical scholars being involved directly to critically conceptualise how to develop social robots into a specific setting in the 2021 HRI conference. However, Tanqueray et al [28] showed that involving different experts from different fields yields different findings, all of which are key to make an appropriate social robot in a given context. This resonates with the issue that developers are strangers in the dataset [10] and potentially mirroring harmful norms [9], since developers have to understand how to navigate a field from their own understanding, due to lack of formal legal norms or other normative guidance, they might be overlooking important social structures. Consequently, developers may reproduce the rhetoric of the institutions who hold more power and legitimacy, which is an issue in itself as it leads an oversimplification of individuals [22] and to reproducing the rhetoric from powerful institutions, as discussed by Ruha Benjamin [20]. This is all the more important as they reflect themselves as well as choosing who to collaborate with, although the community is increasingly aware and critical of this. This shows a need for critical scholars to be part of the collaborations and help inform the design of social robots, and help make difficult normative decisions.

6. Limitations, Future Work and Conclusions

This study represents a screenshot of a sample of the HRI community undertaken in 2021, which was executed fully digitally during a global pandemic, and consequently may not necessarily reflect all developers’ practices. More work needs to be undertaken to test the generalisability of these findings. Further work should also be undertaken in researching assumptions HRI developers base themselves on when deciding to innovate in a new sector. This research, stemming from a SoL approach, would deepen understanding on how social and legal norms are involved.

The study points to three main themes through a socio-legal data feminist lens, which make up the mirroring norms, privilege hazard, and stranger in the dataset: (I) The HRI Community self-regulates itself through informal means, mostly due to the lack of formal legal norms to support developers. This in turn means that (II) they self-rely on one another and reflect not only through their own selves but also through the community. This showcases how (III) developers work towards a common goal to bring social robots in society, although this might include only certain (powerful) institutions. These have major normative impacts on the development of social robots overall and what those
social robots reflect through the way they are developed and programmed. This is somewhat alarming as no explicit legal norms are perceived to steer their decisions, nor do they by default account for the complexity of society. However, the openness of the community shows that critical scholars are likely to become even more active in informing the design of social robots as they are aware of the dataset and the power-related complexities within specific areas. This should be welcomed by critical scholars including lawyers and policymakers, as these robots continue to make headway into society.

Acknowledgements

This work was supported by 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