LUND UNIVERSITY

Behind Every Social Robot Finds Itself a Community of Developers:

A Socio-Legal Exploration on Developers of Humanoid Social Robots with a focus on the Context of Depression Diagnosis and its Embedded Gender Norms.

Author: Laetitia Tanqueray

Lund University
Sociology of Law Department

Master Thesis (SOLM02)
Spring 2020

Supervisor: Amin Parsa

Examiner: Anna Lundberg

ABSTRACT

The rise in new digital technology is bringing with it new ethical questions around how we, as a society, can trust them. However, what if the question was before the technology itself? What if the role of the developers was influencing the way those new technologies were being programmed? This thesis situates itself at the point of exploration, where developers of humanoid social robots look for ways to advance the human-robot interaction as a field as well as getting social robots commercialised. By placing it in the context of depression diagnosis, where seemingly there are gender norms embedded within the diagnosis, this thesis demonstrates how developers might navigate such a sensitive topic. To achieve this goal, interviews with developers and an ethnography of the HRI Conference were undertaken. Through a socio-legal and digital feminist theoretical framework, this thesis pinpoints the huge potential normative consequences due to embedded norms which derive from developers themselves, as well as the data. This thesis concludes with seven policy recommendations to support developers in their exploration.

Keywords: Socially Assistive Robots, HRI, Sociology of Law, Depression Diagnosis, Gender norms

Word count: 21,952

ACKNOWLEDGEMENTS

When I applied to this Sociology of Law Master's, I was certain that I would only ever research young carers, as this topic is very close to my heart. However, as soon as I began the master's, I realised the breadth of research available and that I could use this as my own intellectual playground. For this, I am forever grateful to the amazing Sociology of Law Department for allowing me try out various quirky avenues. Thank you specifically to my supervisor, Amin, for helping map this out at a thesis level. Also thank you to Stefan for believing in my capabilities and showing me how to navigate academia. And of course, I'm very grateful to my peers for enthusiastically receiving my ideas and ensuring that I remained academically rooted. A particular thank you to Carlo, Tamy, Jasmijn and Juliana – the SoL family – the (much) needed sunny family in cloudy Lund.

During my master's I learnt so much about feminism and gender issues. These issues are by default sticky and controversial. I am forever indebted to Juliana, Malin and Lory for enlightening me on such topics and teaching me how to mindfully speak of such issues. I have become a better scholar and person because of it.

Trying to navigate a new country in a pandemic, which both brought many obstacles, as well as undertaking a thesis, was mentally taxing. I have had the chance to go into work, where I have had the most supportive and daily fun-filled conversations with my colleagues. I am also grateful to my amazing friends (virtually and physically) for checking up on me and encouraging me. Marco, I hope I have thanked you enough for your unconditional support throughout this journey, and how much it has meant to me. Mum, Star and Oscar, you're the trio that the whole world should have; you are the best cheerleaders and influencers in my life.

This thesis could not have been possible without my exceptional interviewees, the presenters at the HRI Conference and the HRI community as a whole. Thank you, Ginevra, for acting as my gatekeeper into social robotics; and Rafsan for explaining technical terms. Katie Winkle, Connor McGinn, Eduard Fosch-Villaronga, Jane, Charlie, Karen, and many others: all of you have shaped my thoughts for this thesis, and I hope you can hear your voice clearly throughout. You truly are an inspiration.

Finally, I want to dedicate this thesis to anyone struggling with mental health and people living alongside it. No matter your gender, I believe you. I would have loved to share this thesis and the process of it to someone very dear to me. But for now, I can only hope that one day he will read it.

ABBREVIATIONS

AI	Artificial Intelligence
AIHLEG	High Level Expert Group on AI set up by the European Commission
EU	European Union
GDPR	General Data Protection Regulation
GP	General Practitioner
HCI	Human-Computer Interaction
HRI	Human-Robot Interaction
Q&A	Question and Answer
SAR	Socially Assistive Robot
SoL	Sociology of Law
VH	Virtual Human
WHO	World Health Organisation

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

Section 1: Background of the Study

> "Engineers today hold many responsibilities in their role as developers of new technologies. As anxieties about how technological decisions might play out in society increase, engineers are expected to take into account broader societal contexts, thus going beyond the traditions of requirement specifications and technical development." (IEEE et al., 2021, p.7)

The above quote derives from a hackathon between engineers discussing ethical AI developments in September 2020 (ibid). The context of this hackathon was to bring forward the voices of engineers developing AI tools and their increasing responsibility towards ensuring that these systems are ethical. Within this hackathon, engineers also mentioned the broadness of AI guidelines, rendering them inadequate to guide developers in ensuring they are creating ethical AI (ibid). The excerpt thus demonstrates the increasing pressure on developers, which is beyond the engineering realm; yet, there is still an expectation on engineers to ensure that their creation is ethical as well as to continue innovating new technologies.

"AI Ethics" has become a benchmark over the last few years which companies developing AI systems should align with (see for example, High-Level Expert Group on Artificial Intelligence, 2019). However, this buzz around AI overlooks the workforce and developers' work beyond coding. For example, the Ethics Guidelines by the AIHLEG recognises the workforce, however, it only explicitly mentions that the workforce should use adequate data for AI systems (ibid). Furthermore, guidelines and laws do not necessarily accommodate for the initial exploration needed to create new technologies. Yet the need for exploration is undeniable. Developers have to explore in order to create new technologies, in the hope that the innovation will become widely commercialised and used by the general public. In line with this, this thesis investigates the explorative phase, through a socio-legal and data feminist lens focusing on humanoid SARs in the setting of depression, to demonstrate the potential normative impact developers will have on their innovations, and the lack of guidance from regulations.

Generally, during the exploration phase, developers will have to make important normative decisions which potentially have harmful effects on society. Looking at the data deriving from depression diagnosis, the diagnosis is seemingly skewed: GPs are over 50% of the time likely to misdiagnose depression; women are twice as likely to be diagnosed with depression than their male-counterpart before 55-years-old; and, partly due to Western stereotypes and lack of specific research, men are much less likely to be diagnosed with depression (Carey et al., 2014; Girgus & Yang, 2015; Mitchell et al., 2009; Oliffe et al., 2019; Walther et al., 2021). This should not be underestimated, as WHO has categorised depression as a common disorder with over 264 million people suffering from it in 2018 (2020). Accordingly, the embedded gender norms within depression diagnosis will be part of the data developers will use to automate the screening within SARs. Hence developers of social robots will have to take a normative position on depression diagnosis, even if they do not explicitly intend to. However, due to the lack of guidelines, developers of social robots currently navigate this alone, as this thesis demonstrates. Although it is worth pointing out that despite this lack of support, automating depression screening can be successfully executed: for instance, the HCI field is making headway in screening for depression: simply by evaluating filters on photos posted on Instagram, algorithms detect depression 70% of the time correctly –which is much higher than GPs at present (Islam et al., 2018; Mitchell et al., 2009). In other words, developers will still find successful ways of implementing automation, although it can overlook societal discrimination, which can be harmful to society, especially men in this setting.

In the context of this thesis, HRI developers, concerned with interactions between humans and robots (hence HRI), include engineers as well as data scientists and anyone else that needs to be involved to programme a particular part of the robot. This may include psychologists to understand how the robot should interact with the users in a medical setting. This is key since social robots are robots designed to interact and communicate with humans by replicating accepted social norms (Bartneck & Forlizzi, 2004, p.3; Dautenhahn & Billard, 1999). Social robots are made of hardware which embed AI tools and various other software, which can mimic HCI findings (European Commission, 2020, p.16). Humanoid social robots specifically, appear human-like, either through imitating a human body or a human face (Furhat Robotics, 2021; SoftBank Robotics, 2021; see figure 1&2). However, social robots are not yet widely commercialised. Nevertheless, innovations and advancements in humanoid social robots suggest that they could be used to help screen for diagnose depression; these types of robots, that are used to assist humans, are referred to as SARs (Fosch-Villaronga & Albo-Canals, 2019, p.82). Accordingly, this thesis will interchange between SARs, social robots, robots and HRI. This is done intentionally unless specifically stated that the findings are only related to SARs.

Since this study attempts to shed a light on HRI developers' exploration, and how this might affect their exploration around depression, it is worth mentioning social structures. Social structures in the context of this thesis refers to systems, albeit institutions but also relations between people, which are dependent on a breadth of factors. In this instance, the social structures are primarily focused on gender norms, whereby on the basis of the person's sex, the medical institution—made up of medical professionals— will diagnose differently (Criado Perez, 2020; Keller, 1987). This can potentially be transferred and amplified into SARs if the developers do not reflect on and take into account social structures (Benjamin, 2019; D'Ignazio & Klein, 2020; Larsson, 2019).

The interest of this thesis lies in understanding developers' practices and methods of exploration to advance the field of social robotics, and its normative consequences. To do so, this thesis relies on two socio-legal theories: Larsson's conceptualisation of normative mirror effect as well as Hydén's algo norms – both well-established socio-legal scholars (Hydén, 2020; Larsson, 2019); as well as

D'Ignazio and Klein's feminist approach and their concept of Strangers in the Dataset (2020). To achieve this, interviews with developers in humanoid social robots' manufacturing companies have been undertaken as well as attending the week-long HRI Conference 2021 to conduct an ethnography on HRI developers.

Section 2: Research Questions

The overarching research question is the following:

> How is the role of HRI developers mirrored into humanoid socially assistive robots and how might this impact their exploration into depression diagnosis, especially regarding the embedded gender norms?

The sub-questions are as follow:

1. How are HRI developers advancing their field?
2. How are current regulations regarded by developers to develop and design socially assistive robots?

Section 3: Objective of the Study

The objective is to demonstrate the normative power developers have: they are not merely innovating; they are also having to choose which norms to reproduce when developing the robots. Accordingly, this thesis primarily aims to bring forth the developers of social robot's experiences as well as their understanding of their current role in society. The setting of depression diagnosis demonstrates what developers should be aware of and how they might accommodate for it.

Developers do not have specific codes of conduct to follow or many regulations to lean on whilst exploring. In order to demonstrate this, this thesis focuses specifically on the possibility of advancing social robots in the realm of medicine, by creating SARs to help screen for depression. The interest stems from the already existing societal inequality in medicine, focusing on gender norms, in depression diagnosis. The importance lies in demonstrating the burden on developers to innovate in areas loaded with social inequalities, where due to their lack of expertise in this, they may reproduce them. This should not be underestimated as SARs could be scaled up to diagnose many individuals who might be experiencing depression. In my conclusion, I propose seven policy recommendations in order to support developers in taking normative positions.

Section 4: Delimitation of the Study

This thesis focuses on HRI developers: how they innovate and advance social robotics. This thesis ties the development of social robots to depression diagnosis, although there are currently no commercialised SARs able to do so or used for that purpose. Thus, the study demonstrates how developers, as a community and individually, explore new avenues for social robotics; as well as the communal norms between developers, which will impact how they programme SARs to screen for depression, and impact the issues around embedded gender norms.

This thesis bases itself on the Western medical practice to recognise depression. Thus, the literature reviewed on depression is placed in this Western culture (i.e., from Europe, Australia and the USA). This study itself is further delimited, since it relies mostly on presentations and interviewees based within the EU. Accordingly, although social robots are making advances throughout the world, especially high-income countries, this thesis concentrates on the automation of SARs in a Eurocentric context.

Gender is also important to delimit here. Since the thesis concentrates on developers, which are made up of mostly engineers, the notion of gender and gender norms are still at an early stage and solutions are still highly debated. This results in this thesis understanding gender in an essentialist way, meaning in the simplistic biological binary differentiation of sexes (i.e. man and woman). This was intentional despite myself being an ally to the LGBTQ+ community and hoping to take a more advanced stance on gender. However, for now gender issues in AI as well as research in diagnosing depression, are still based on the basic sex binary. Attempting to rectify this would require more than one master's thesis.

Section 5: Structure of the Thesis

This thesis is divided into five main chapters and a conclusion. Chapter 2 defines key concepts which are relevant to this thesis. Chapter 3 reviews extensively the literature surrounding issues of depression diagnosis and new digital technology. Chapter 4 discusses the theoretical framework which veers this study. Chapter 5 explains the methodology to answer the research questions. Chapter 6 presents the empirical findings. Finally, chapter 7 concludes this thesis.

CHAPTER 2: DEFINITIONS OF KEY CONCEPTS

This chapter is specifically designed to enhance the reader's knowledge on social robots, SARs and AI (section 1). Section 2 defines depression and its diagnosis, and the final section points out relevant regulations (section 3).

Section 1: Defining Social Robots and Socially Assistive Robots

The term "robot" was coined by Karel Čapek in his 1920s play, R.U.R. (Capek, 2004). In its original Czech form, "robot" means "forced labour". This demonstrates the characteristics of a robot: it automates labour that humans are deemed capable of undertaking (Moravec, 2021). Social robots advance this traditional view of robots since they are able interact in a sociable manner with humans; in contrast, traditional robots are usually anti-social and thus cannot communicate with humans.

There are various definitions of social robots. For the purpose of this thesis, they are understood as physical machines with distinctive personality and character, which perceive and express emotions as well as communicate through the use natural cues, such as gaze and gestures (Fong et al., 2003, p.145; Fosch-Villaronga et al., 2020, p.443). Accordingly, through various software, which can include AI, the social robot can interact and communicate with humans for an intended purpose— which is achieved through cameras and sensors (Bartneck & Forlizzi, 2004; Mokhtar, 2019). In brief, the robot will be able to recognise and engage "with humans by following the behavioural norms expected by the people with whom the robot is intended to interact" (Bartneck & Forlizzi, 2004, p.3; Dautenhahn & Billard, 1999, see also figures 1&2). As previously mentioned, for the most part, SARs follow the same programming as social robots, however SARs are created specifically to assist users, such as screening for depression.

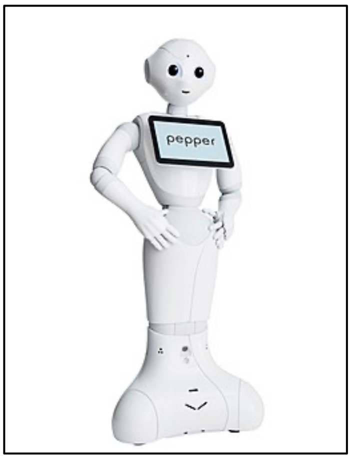


Figure 2: Pepper Robot, a humanoid social robot. Softbank Robotics Europe, CC BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0>, via Wikimedia Commons (2021).

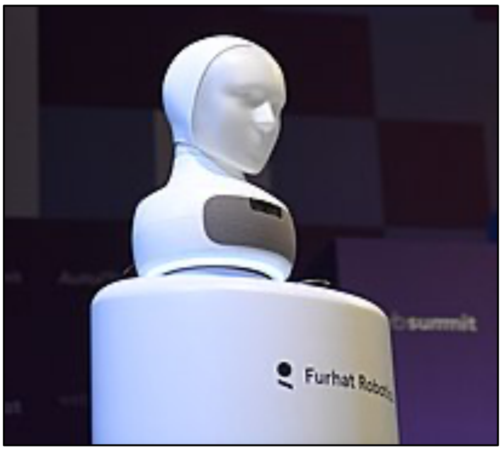


Figure 1: Furhat Robot, a humanoid social robot. Web Summit, CC BY 2.0 <https://creativecommons.org/licenses/by/2.0>, via Wikimedia Commons (2021).

Although the main focus is on social robots, I also refer to AI in this thesis. AI allows the robots to autonomously make a decision which will be based on a certain algorithm. AI alone as a term is very generic; "AI" usually refers to the scientific discipline of AI, which has been around since the 1950s. When referring to AI being used in a specific application, it is either an AI tool/system which can encompass machine learning or deep learning, however the two latter do not need to be outlined for this thesis. According to the AIHLEG:

> "AI systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions." (High-Level Expert Group on Artificial Intelligence, 2019, p.36)

This definition demonstrates that the system/tool requires a level of autonomy and adaptivity in a given context and relies on datasets. The AI tools require a lot of data in order to function. This data will make up a dataset, which the AI tool will be trained on, and over time learn from and find patterns to repeat. Although AI systems rely on the human workforce in the first place to gather the data and then learn from it. However, data is also required for more traditional software, so that the developers know what they should programme based on the data they have. Thus, data is vital, including outside of AI systems.

Turning to social robots and AI, AI makes up part of social robots, but social robots does not only function on AI systems. This can be seen through the diversity of teams within social robotics. The physical appearance of a robot will be made by, usually, the agility team. Teams specialised in interaction and emotions, who train the way the robot will 'decide', for example, when to wave to somebody. In the instance of waving this can rely on AI, where the robot will be trained through reinforcement learning: if the robot does it at the right time it will be positively rewarded. In the instance of telling a joke, it is likely to be done through supervised learning, whereby the robot will detect a sad emotion and choose in the context to make a pre-set joke. Thus the pre-set joke will not be generated by AI.

Section 2: Depression and Diagnosis

It is important to appreciate the intricacies of depression and what it means to be diagnosed with depression. Here I do not discuss the stigmitisation of mental health, but the way depression might affect individuals and be diagnosed. This thesis focuses specifically on depression that medical doctors will need to diagnose and treat; accordingly, when using 'depression' in this thesis, it refers to clinical depression or also known as major depressive disorder (APA, 2020; NHS, 2019b). Consequently, to diagnose this type of depression requires a screening by the patient's GP to rule out any other possible medical conditions (NHS, 2019a). After this, it is usual for a trained psychologists or psychiatrist to evaluate the patient and diagnose the patient.

Depression symptoms vary from individuals, however it will negatively affect the way the individual feels and acts (APA, 2020). Depression should be seen on a spectrum from mild to severe (NHS, 2019c). The symptoms may vary from emotional to physical to social issues, which can range from thoughts of death to slowed movement to avoiding contact with friends (APA, 2020; NHS, 2019c).

When diagnosing depression, it is likely that the doctor in charge will evaluate through an interview and physical examination (APA, 2020). The interview is likely to be made up of a questionnaire, although there is no standard questionnaire. Once the diagnosis is reached, the patient may undergo a series of therapy sessions and/or medication to help her cope with her condition.

Section 3: A Note on Regulations

As a socio-legal thesis, it is worth mentioning some laws, standards and guidelines which I merge together and call regulations.

The most relevant enforceable law is the GDPR which is a key law in this area (2016). The GDPR was drafted by the EU, and concerns the processing of EU citizen's personal data. The GDPR has been referred to as "the law of everything", since its scope is to try and tackle issues digitisation brings (Purtova, 2018). The GDPR will still apply to developers when they are exploring since they will still rely on data –which might include personal data from the EU. There are other obligations on data controllers, but these are the most important to bear in mind for this thesis

Standards will be set by specific organisations such as the ISO (International Organisation for Standardisation) (ISO, 2021); or the IEEE – the "world's largest technical professional organization dedicated to advancing technology for the benefit of humanity" (IEEE, 2021). Most of the standards emanating from these organisations focus on very particular parts of the robot.

There has been a growing popularity around the notion of "ethical AI" as shown in the introduction. This can be viewed by, for example, the IEEE's Global Initiative on Ethics of Autonomous and Intelligent Systems (2019) as well as the AIHLEG which drafted Ethics Guidelines for Trustworthy AI (2019). Neither put much emphasis on the role of the developers, instead their emphasis is placed on what the objectives of AI tools should be. Although, the AIHLEG has been used as a basis for a proposal by the EU on an enforceable Regulation on AI (European Commission, 2021).

CHAPTER 3: LITERATURE REVIEW

Conducting a social science literature review from a continuously innovating and nascent STEM field proved to be difficult. The traditional way of writing key terms in well-known academic search engines (e.g. EBSCO and Web of Science) were mostly fruitless. Consequently, for the literature review, a snowball review method was utilised. Meaning that through articles I found – either from some searches on EBSCO, the HRI Conference I attended or in general academic books— I was able to find more specific and helpful literature.

The literature review is five-fold: section 1 looks at depression diagnosis currently without the use of technology; section 2 looks at new technology and depression; section 3 spotlights social structures and norms within technology, and section 4 focuses on developers themselves. Finally, section 5 merges the findings from the previous sections.

Section 1: The Current State of Depression Diagnosis

1.1: Depression Diagnosis in Research

Historically, healthcare has been a male-dominated field where there are various forms of gender biases (Keller, 1987). Keller often describes this as sex being the obstacle to gender, just as science can be an obstacle to nature: in the sense that objective science is unachievable if done by a specific group of people (usually white men) (ibid). This applies to sex too; sex does not make up the entire person's gender. Keller demonstrates how this can also affect the researcher and medical professionals themselves: they will act according to the norm (in this case male dominated) in order to be accepted by their community (ibid); in other words, female doctors might reproduce the male-dominated practices. This perpetually reproduces biases and discriminations which may not reflect the actual medical condition the person is facing. Instead, as Keller explains, science should be made by every person qualified to do so for every type of human (ibid).

Criado Perez's highly-acclaimed book, *Invisible women*, tackles issues around women being silenced in society due to gender – the social meaning ascribed to the female body – and not sex (Criado Perez, 2020, p.XIV). In medicine this transpires through what is referred to as the 'yentl syndrome'; whereby "women are misdiagnosed or poorly treated unless their symptoms or diseases conform to that of men" (Criado Perez, 2020, p.217).

With regards to depression diagnosis, the 'yentl sydrome' has two consequences. The first is that diagnostic tests are developed around male bodies, even where women are more at risk (ibid, p.220). Thus, women are claimed to be 70% more likely to suffer depression than men, yet animal studies on brain disorders are five times more likely to be done on male animals (ibid, p.205). The second consequence, which leads on from the first, is that women are less likely to be believed that they are in pain and instead are labelled as mad (ibid, p.225). Both of these consequences echo Keller's point around the medical institution (1987). Accordingly, women are two and a half times more likely to be on antidepressants than men (ibid, p.226). Criado Perez suggests that this disparity in antidepressant prescriptions may be that women in physical pain are dismissed as being emotional (ibid).

Although Criado Perez's overall theme is about women being invisible in data, depression in women is well represented (ibid). In other words, depression diagnosis is well established for women, despite research generally overlooking the female anatomy. This is demonstrated by Ballantyne and Rogers, who conducted a study to determine the proportion of female participants within research studies used to develop national clinical guidelines (Ballantyne & Rogers, 2011). They found that the research studies in question, which focused specifically on depression, used mostly female participants (66% were female participants) (ibid). Put differently, the research studies used for policy making was based on data which was mostly represented by women. Ballantyne and Rogers conclude that this "accurately reflects the population affected by this condition" since more women are deemed to suffer from depression (ibid, p.1303). Subsequently, the question is not so much about women and depression – this is extremely well researched (see for example Atwood, 2001; Bacigalupe & Martín, 2021; Brommelhoff et al., 2004; Rippon, 2019) —the question lies more with the lack of diagnosis in men.

A much smaller proportion of research is dedicated to understanding the reason why men might not be as diagnosed with depression than their female counterpart. A German study aimed to understand if a certain medical examination for depression diagnosis was actually allowing men to report their mental health struggles (Walther et al., 2021). The issue, the researchers point out, is that there is a low uptake on depression treatment and thus low diagnostic reliability for men (ibid, p.1; Oliffe et al., 2019). Walther et al argue that masculine gender norms may play a role, such as "restrictive emotionality" which is entrenched in Western culture (ibid, p.2). These gender norms mean that men conforming to traditional masculinity might externalise their depressive symptoms through anger or substance abuse (ibid). This leads to undiagnosed depression in men, since those signs are not necessarily viewed as depressive symptoms (ibid).

1.2: Depression Diagnosis in Practice

It is worth mentioning the use of questionnaires for depression diagnosis. A healthline article, reviewed by a speciliased doctor, presented various depression questionnaires (Healthline Editorial Team, 2018). This articles demonstrates the lack of aninimity in diagnosing depression through the various questionaires available.

Turning to the medical staff, Mitchell et al conducted a meta-analysis study on clinical diagnosis of depression in primary care, focusing on GPs diagnosing depression (Mitchell et al., 2009). This was justified as GPs being the first port of call for patients seeking help for their mental health. Accordingly, Mitchell et al reviewed 41 studies, which included 50,371 patients, to demonstrate the difficulties in diagnosing depression by GPs. GPs correctly ruled out depression most of the time, however they identify depression in 47.3% of patients with depression; meaning that over 50% of the time, GPs misidentify cases of depression (ibid, p.616). GPs were more likely to diagnose depression accurately when using specific scales rather than their own judgement, although, Mitchel et al claimed, this is partly due to GPs being overcautious in diagnosing depression (ibid, p.617). Mitchell et al use their findings to raise awareness and recommend more resources for GPs as well as more time with patients to screen and conduct a second assessment on potential patients with depression (ibid).

A similar study conducted in 2014 on Australian GPs found similar results (Carey et al., 2014). Carey et al also mention the impact of stereotypes, leading to GPs over-diagnosing women with depression (ibid, p.572). Accordingly, using a standardised depression tool could improve depression diagnosis, although this requires more resources (ibid, p.576). Carey et al also note that the use of a mobile computer tablet for patients to self-assess could "alleviate several barriers to standard screening procedures" (ibid).

Section 2: Technology and Depression Diagnosis

2.1: Human-Computer Interaction and Depression Diagnosis

HCI as a field is much more advanced than HRI. It is therefore useful to showcase some of the advances in the HCI field, in order to demonstrate HRI's influence and where it can be heading towards in the context of depression diagnosis.

Topol, a world-renowned cardiologist and medical researcher, displays the use of AI in medicine, and how the appropriate use of AI will benefit both patients and medical staff (Topol, 2019). Topol's main aim in *Deep Medicine* is to demonstrate that "as machines get smarter, and take on suitable tasks, humans might actually find it easier to be humane" (ibid, p.4). Consequently, in the chapter dedicated to mental health, Topol showcases the great strides being made in HCI to enable earlier diagnosis, more support for people as well as a more accurate overview of the individual patient (Topol, 2019, Chapter 8).

Research indeed is pointing towards that direction, although it is demonstrating that people might even prefer talking about their mental health to virtual agents. Lucas et al, through the use of a virtual human (VH), demonstrated that people were more likely to disclose to VH (Lucas et al., 2014). The VH interviewed participants through a computer screen. The groups were split into two, one group was told that the VH was a computer (not human), whilst the other was told the VH was controlled remotely by a human. Over time, the VH would ask more and more intimate questions. Lucas et al found that the participants were much more willing to disclose, especially expressing sadness, to virtual humans/computers (2014).

These types of findings have led to the development of virtual support, such as apps. One, named ‘Woebot’, was developed by engineers and psychiatrists and is used by hundreds every month (Woebot, 2021). Woebot uses advanced AI to understand and adapt to human language and answer accordingly to each person. This scaling up should not be underestimated: through chatbots using AI in a mental health support context, users are able to get appropriate support. Furthermore, the number of users Woebot had in its first few months was more than a psychologist alone could see in a hundred years (Topol, 2019, p.167).

Aside from specific apps which is used by people seeking this type of help, AI tools are also being developed for depression detection on social media platforms. This involves all individuals that use social media platforms, instead of downloading a specific app. For example, through the use of filters on photos on Instagram, the AI tool accurately detected depression in 70% of cases (Reece & Danforth, 2017). On Facebook, Islam et al., used emotion detection techniques coded in an AI tool to detect depression in users comments (Islam et al., 2018). The tool was accurate 60% to 80% of the time. Both of these are successes when comparing to GPs 50% accuracy in detecting depression (Mitchell et al., 2009).

2.2: Human-Robot Interaction and Depression

Social robotics as a field is at early stages, however developments are fast-paced, which has included mental health and wellbeing (Scoglio et al., 2019). Scoglio et al investigated, through the means of a systematic review, the current literature on SARs in mental health intervention (ibid). Scoglio et al motivates their research by explaining that there is a disparity between people getting mental support which could be alleviated through the use of technology such as telehealth/mobile health delivery methods (ibid, p.1). However, there is low uptake by clinicians and users as well as low engagement from the users. Nevertheless, Scoglio et al argue that SARs seem to be well suited to mental health and well-being support. This is due to SARs' uniqueness in being able to be a platform for intervention as well as being an intervention in their own means (ibid, p.2). Said differently, the SARs can use a pre-set programme – not just intended to be used for robots— to act as an interface, or they can intervene directly to provide support. This means that SARs "can learn and engage socially with individuals while also presenting interventions to users similar to mobile apps" (ibid). Accordingly, SARs interact and connect with users whilst also integrating various apps to provide various forms of mental health support for users. However, Scoglio found only 12 relevant articles in this field, all of which did not include people who were diagnosed with mental health issues. Furthermore, the studies mostly concentrated on elderly (seven out of twelve), one concentrated specifically on women aged 19-45 years old and two on medical staff (ibid, p.4). Humanoid SARs specifically mostly had positive outcomes: users felt more relaxed or in a better mood and, medical staff found SARs helpful (ibid, p.5). This was also found in a different scoping review, whereby depressive symptoms were alleviated through the use of SARs within elderly care (Abdi et al., 2018).

Although it is clear from the Scoglio et al's research, SARs are at an early stage, it is worth pointing that automation of mental health support seems to be mirroring current trends. The studies either concentrated on the elder population or only women in the setting of mental health issues for SARs. However, another scoping review focused only on SARs for elderly care, reported that out of 1574 participants, 71% of them were women (Abdi et al., 2018). Abdi et al point out that this shows gender bias, although to them it was concerning because both genders may view the SARs differently (ibid, p.17).

Section 3: The Importance of Sociology and Law in New Technology

3.1: The Hidden Social Structures Behind New Digital Technology

Benjamin, an Associate Professor of African American Studies and author of *Race After Technology*, uses the "New Jim Code" as a theoretical framework to examine how racism is maintained or perpetuated through "technical fixes to social problems" (2019, p.48). Jim Crow Codes were originally used to upkeep White Supremacy, thus the *New* Jim Code is an updated version, whereby "tech fixes often hide, speed up, and even deepen discrimination, while appearing to be neutral or benevolent when compared to the racism of a previous era" (ibid, p.8). Accordingly, terms such as "progress" and "objectivity" should be used carefully since they weaponise against those who suffer most in already existing oppressive systems (ibid).

The New Jim Code acknowledges that technology attempts to ignore social divisions, which causes to reproduce them as well as to fix social biases "but end up doing the opposite" (ibid). However, Benjamin points out that these are "outcomes" and not "beliefs", since the reproduction of such structures may not be the intention of the developers (ibid, p.17). Instead, Benjamin rationalises it as a move towards privatisation, whereby "efforts to cut costs and maximize profits, often at the expense of other human needs, is a guiding rationale for public and private sectors alike", this automation appears to remove the burden "from gatekeepers, who may be too overworked or too biased to make sound judgments" (ibid, p.30). Consequently, overlooking social divisions to remove the burden from gatekeepers through the use of technology is unlikely to produce an appropriate model for the new technologies.

D’Ignazio and Klein, both software developers and now associate professors, co-authored *Data Feminism* to bring to light the need for feminism in data science. The main argument is based around seven principles in order to demonstrate how developers need to be aware of the data collected, or else the status quo may be maintained (D’Ignazio & Klein, 2020, p.17). The seven principles are therefore (1) examining power, (2) challenging power (3) elevating emotion and embodiment (4) rethinking binaries and hierarchies (5) embracing pluralism (6) considering context and (7) making labour visible (D’Ignazio & Klein, 2020).

Fundamentally, D’Ignazio and Klein explain that there is a need to look at the role of institutions themselves, especially in relation to power structures (ibid). The institution from which the data emanates from needs to be transformed; especially since those are the ones to “produce and reproduce those biased outcomes in the first place” (ibid, p.32). Especially as groups, such as women, are often excessively surveilled when powerful institutions benefit from it, accordingly there may be more data on such groups (ibid, p.39). This was demonstrated in the first section of this chapter in relation to depression diagnosis.

Past data, which makes up the model, is therefore never “raw” but reflective of current social inequities (ibid, p.55). Consequently, issues that are solved through technological solutions are based on data which will benefit the institutions rather than the individuals who are directly impacted (ibid, p.40). This lack of raw data is key, since this demonstrates that more data will not solve the issue.

Conveying this issue to women, often the narrative around women is that they require protection and have no agency (ibid, p.59). Accordingly, the ‘raw’ data collected will often supplement this narrative (ibid). Benjamin describes this as “allure of objectivity” whereby false beliefs are viewed as objective and thus self-reproduced (Benjamin, 2019, p.53). This is worsened by the approaches to tech developed claimed to be “colour-blind, gender-neutral, and class-avoidant” (ibid, p.63).

3.2: Sociology of Law and Social Structures in New Digital Technologies

Socio-legal scholar Larsson, in his article "The Socio-Legal Relevance of Artificial Intelligence", gives a broad overview of the contemporary everyday versions of AI and automated decision-making tools (AI is part of the latter tools). Larsson aims to demonstrate the need to include society as a whole in order to reflect on what norms the AI tools should be trained on (2019). SoL is needed to understand the normative perspective within these types of tools. However, it is important to note that according to Larsson, the current laws are well-established and address issues of discrimination, markets and data protection (ibid, p.591). The challenge is beyond regulation and SoL pinpoints how to address the translation of societal structures and values within the AI systems. Accordingly, Larsson's article is framed around Fairness, Accountability and Transparency (FAccT) to enable a socio-legal discussion where he coins a concept named "mirroring norms" (ibid, p.589).

To establish mirroring norms, Larsson draws upon Ehrlich and Petrazyki's theories, two Founding Fathers of SoL, to reiterate that there are formal laws as well as living law (Ehrlich) or intuitive law (Petrazyki) to regulate society. Both theories enable a more empirically based approach to law which goes far beyond the realm of State/formal law (ibid, p.589). This duality of laws is fundamental to Larsson to demonstrate the mirroring of norms. The informal laws make up a diversity of norms according to different groups of people and context. Those norms are important since they make up the data which developers will use and the AI tools will learn from. However, usually which norms are reflected is unknown, especially at the learning stage (ibid).

During the learning stage, also referred to as the black box, the AI tool will learn based on the data it has been fed by developers. The developers devise this data and define their desired outcome. Accordingly, the AI tool will have to choose which norms to reproduce in order to get the desired outcome, which can directly conflict at a micro level (between informal social norms) and at a macro level (between legal norms and social norms). This results in the AI tools reflecting –mirroring norms— by reproducing the social structures as well as potentially amplifying them (ibid).

Hydén, another socio-legal scholar, focuses on advancing SoL research in new digital technology, particularly AI. According to Hydén, there is a need for SoL to focus beyond the law in order to understand the normative consequences of new technologies due to their algorithms, hence he coined the term "algo norms" (Hydén, 2020). Hydén uses Lessig's theory "code is law" to contrast between code and algorithms. Lessig points out that code is constrained by four forces: law, social norms, market and architecture (Lessing, 1999, in Hydén, 2020, p.360). Consequently, the codes establishing digital technologies are potential regulators and the code writers themselves, as Hydén articulates, become "responsible for social construction" (ibid).

Accordingly, Hydén formulates that if code is law, then algorithms are norms (ibid). Hydén does not define code, but defines algorithms as "consist[ing] of what to do, with what and in what order" (ibid, p.361). Whilst "code is law" concentrates more on understanding how to regulate such technologies, "algo norms" are the indirect effects of the technologies themselves. This distinction is important, as Hydén does not veer towards the law, instead the focus is on the normative consequences embedded in new digital technologies (ibid, p.363).

Section 4: Developers of New Digital Technology

4.1: Who are Developers?

As a whole, the tech industry is renowned for a lack of diversity in its workforce. Benjamin refers to the tech labour force as "deeply unequal across racial and gender lines" (2019, p.58). This may lead to blind spots within the development process (Larsson, 2020, p.581). UNESCO produced a report outlining the persistence and severity of the gender gap in digital skills (UNESCO, 2019). All these result in a homogenous workforce which will possibly reproduce gender stereotypes. This is well documented with regards to voice assistants, such as Alexa, which are female voices, polite and subservient – in ‘accordance’ with women’s virtues (ibid, p.98). Thus, generally developers tend to be from a certain gender as well as a certain class.

However, this view of the tech industry is not necessarily mirrored on developers as individuals. Developers who handle data for the purpose of automation have been defined as follow:

> “[P]eople who work with data are alternately called *unicorns* (because they are rare and have special skills), *wizards* (because they can do magic), *ninjas* (because they execute complicated, expert moves), *rock stars* (because they outperform others), and *janitors* (because they clean messy data)” (D’Ignazio & Klein, 2020, p.133).

Accordingly, these people referred to as developers in this thesis, are seen as experts with unique skills. They make up a group with privilege, who may play a role in “upholding oppressive systems”, especially if they do not reflect on their position (ibid, p.63-64). Though, Benjamin argues that developers can be reflective, but they might not individually have the power to intervene against the company they work for (Benjamin, 2019, p.61). There are also other powers at play, such as within the company or research group — the hierarchy between co-workers and the team they are part of – since engineers will unlikely work alone on an entire project (IEEE et al, 2021, p.18). They are also constrained by commercial obligations, such as non-disclosure agreements (ibid, p.24).

Developers also tend to fall under, what D’Ignazo and Klein call, “privilege hazard”: “those who occupy the most privileged positions among us—those with good educations, respected credentials, and professional accolades— [are] so poorly equipped to recognise instances of oppression in the world” (D’Ignazo & Klein, 2020, p.29). Thus, these small groups might not recognise instances of oppression, yet they are shaping the data to then scale up to users around the globe (ibid, p.28).

4.2: Expectations on Developers

On an individual level, D'Ignazio and Klein point out that developers view steps such as cleaning their collected data, to be able to use a functioning model, as merely "technical conundrums" (2020, p.65). Meaning that social context, values or even politics of data is not usually reviewed nor reflected upon (ibid). However, these "technical conundrums" will make up the classification systems in order for the developers to have a working infrastructure for the algorithm (ibid). These infrastructures may become naturalised and normalised without reflection on the systems until issues arise (ibid, p.104). According to Larsson, those moments make developers take "normative position on issues they would prefer to avoid", which could unintendedly reproduce societal issues (ibid, p.590).

To Larsson, there are two questions about developers: "should they [developers] reproduce the world in its current state or as we would prefer the world to be? And who gets to decide which future is more desirable?" (ibid). This issue has been recognised by engineers themselves, at a hackathon held to voice current challenges and struggles they are facing when developing new technologies using AI. These engineers view their role as *de facto* impacting society (IEEE et al., 2021, p.3). Consequently, the engineers at the hackathon "readily acknowledged that they must make choices in the development of AI technologies and that these choices can have ethical implications" (ibid, p.11). Thus, it seems that engineers' technical expertise is "no longer enough" as they have to take into consideration broader societal issues (ibid, p.19). This results in an expectation on engineers to make important normative choices by themselves.

Expectations also arise from guidelines, such as the AIHLEG's Ethics Guidelines, however engineers point out that these "place a fair amount of responsibility for how AI systems are designed and developed on engineers" (ibid, p.32). However, these guidelines are mostly unused by engineers to make choices during the development process (ibid, p.32). This feeling has been echoed by developers within social robotics used for therapy purposes (Fosch-Villaronga et al., 2020). Fosch-Villaronga, Luts and Tamò-Larrieux held four international workshops with experts to discuss robots' ethical, legal and societal (ELS) concerns (ibid). Experts that attended the workshops felt that they only had guidelines to follow, which are good for innovation, but they are not concrete or enforceable to know exactly what is expected of developers (ibid, p.447). Furthermore, they pointed out that policymakers discuss implications of robots in general but not in a specific context, such as therapy (ibid, p.447-448).

4.3: Developers' Priorities

Regarding developers of social robots specifically, Šabanović —a social roboticist herself—interviewed and observed Japanese researchers to shine a light on the culture in social robotics over there (Šabanović, 2014). Šabanović indicates that Japan as a whole aims to include robots within everyday life, thus researchers are expected to reproduce conservative social values so that social robots will be accepted by consumers. Accordingly, anything beyond the "assumed cultural homogeneity" is viewed as a threat (ibid, p.358). However, researchers rely on their own "cultural standpoint" to justify design choices and modelling "appropriate attitudes toward robotics technology" (ibid, p.359). These bring forth two issues according to Šabanović: the first is the lack of reflection from researchers themselves who do not question nor test the cultural assumptions they make; the second is that social robots are not yet widely commercialised and thus consumers do not contribute to social robots (ibid, pp.360-361).

Similarly, engineers at the hackathon pointed out, commercial concerns are usually prioritised over ethical considerations due to allocation of time and resources (IEEE et al, p.29). Thus, engineers have to prioritise the company's interest, which may also be part of the "privilege hazard" especially if they are to prioritise what they deem to be 'cultural values' (Šabanović, 2014; D'Ignazio & Klein, 2020).

4.4: Developers' view on Depression

Fosch-Villaronga, Luts and Tamò-Larrieux's workshops also included SARs in a depression setting (Fosch-Villaronga et al, 2020). The experts were ELS specialists and engineers, as well as psychologist and cognitive scientists. They disagreed on the use of SARs in a depression setting (Fosch-Villaronga et al, 2020, p.450). One of the experts argued that social robots could reduce the burden of caregivers' workload, whilst another explained how the introduction of social robots to cognitive therapies could lead to "the substitution of human narrative therapists on the one hand and the exacerbation of patient issues on the other hand" (ibid). It would seem that in response, some participants pointed out that the use of robots in therapy "implies in-depth cooperation between engineers and scientists in that particular field" (ibid, p.451). However, these types of discussions between experts demonstrates how developers will have different views on the same issues. These views will result in various normative decisions, such as who they bring on their team to develop and what they aim to tackle: it may be reducing the burden of caregivers or be centred around the patients.

4.5: Developers Innovating Beyond Their Field

Developers' skills are sought after because they can clean up data in order to automate processes— hence their status of rockstars, ninjas, unicorns and wizards (D'Ignazio & Klein, 2020, p.133). However, a direct consequence is that developers are "strangers in the dataset" (ibid, p.130). Thus, whilst domain experts – such as doctors— are able to intuitively understand data and use it appropriately through their training; by contrast, developers are removed from the data that they are automating (ibid). As a result, the data may be available, but the data is not necessarily attainably understood by outsiders who are the developers (ibid, p.133).

Topol has argued that this can have a negative impact on diagnosing depression, since it is for the developers to classify the possible symptoms of depression (2019, p.172). According to Topol, this is "tricky because historically mental health disorders have largely been defined by subjective and clinical features" (ibid). Consequently, being a "stranger in the dataset" can make it hard for developers themselves to reflect on the dataset collected and the consequence of the 'raw' data, as demonstrated in relation to depression diagnosis.

Section 5: Merging the Literature, Showing the Gap

In brief, section 1 has demonstrated the emphasis on women regarding depression diagnosis generally, leaving men out from a lot of research in this area. This leads to women being well represented in studies representing, and over 65% participants in studies used to draw up policies and guidelines. Furthermore, the literature has demonstrated the overbearing burden on GPs to diagnose depression, which they do not correctly diagnose over 50% of the time, which is also impaired by gender stereotypes. Section 2 pointed to the huge leaps made in automating screening for depression in both HCI and HRI fields, with the HRI field unintentionally focusing on women in therapy settings (over 70% were female participants). Section 3 moved away from medicine to look more deeply into the importance of considering contexts of the data. These include social structures from which the data derives from and which norms might be affecting the model—such as the narrative that women need to be protected. Finally, section 4 focused on developers generally and how they are usually removed from the data they are automating.

Overall, the literature points to the need to be critical of historically trusted institutions, such as medicine. There is clearly a need to reflect on the institutions from which the data derives from. There is also a need for the developers themselves to be aware of such practices to be able to automate processes in a way which members of society will benefit from, instead of the paying actors. Seemingly, the ethics guidelines tend to be overlooked by developers although those guidelines also seem to overlook the pivotal roles of developers.

This extensive literature review has demonstrated that developers are constantly innovating in HRI and critical scholars are aware of their normative power.

However, what continuously lacks from the literature is the focus on developers in one specific field. D'Ignazo and Klein speak on behalf of their own experience as software developers to demonstrate how developers can be more reflective as well as more aware of power structures; and Šabanović uses her own expertise to examine the impact of japanese culture on social roboticists. Nevertheless, the literature does not showcase specific developers to show how HRI is moving forward as a community and how individuals are continuously developing to commercialise socially assistive robots. Accordingly this thesis attempts to bring forth the experiences of developers from HRI only. This is achieved by looking at the community generally through attending the HRI Conference, and interviewing some developers from the most advanced social robotics companies. This showcases whether developers of SARs in the setting of depression may replicate and perpetuate the current status quo.

CHAPTER 4: THEORETICAL FRAMEWORK

This chapter focuses on the theoretical framework used to analyse part of my empirical data. As demonstrated in the literature review, social robotics as a field is nascent. As a consequence, there is not much research on the field from a social scientific perspective, including SoL. Accordingly, the theoretical framework I use is mostly made up of SoL scholars: Larsson's concept of "mirroring norms" (2019), as well as Hydén's concept of "algo norms" (2020). To supplement these, D'Ignazio and Klein's concept "stranger in the dataset" to allow a specific focus directly on developers beyond AI tools (2020).

Section 1: Larsson's Socio-Legal Concept of Mirroring Norms

Mirroring norms refers to the data which autonomous technologies rely on to learn from (Larsson, 2019). Since this data is derived from society, it includes "the balanced sides of humanity" but also the "biased, skewed and discriminatory" sides of society (Larsson, 2019, p.575). From this data, the tool replicates and amplifies existing norms; resulting in norms embedded in the data to be mirrored, hence "mirroring norms". This mirroring of norms has huge normative implications for developers, who may unintendedly reproduce biased and discriminatory norms (ibid, p.575;589). Larsson states that there are no quick fixes to this and developers have to take normative positions, even if they would rather prefer to avoid them (ibid, p.589). Accordingly, the challenge is beyond regulation and there is a need to understand how to address the translation of societal structure and values within the AI systems (Larsson, 2019, p.591).

There are two questions Larsson points to regarding developers: "should they [developers] reproduce the world in its current state or as we would prefer the world to be? And who gets to decide which future is more desirable?" (ibid). These questions enable around developers and the HRI community generally. Mirroring norms, as a concept, accommodates for the intricacies of exploring new technology in sensitive areas, such as depression, and the default normative position-taking by developers.

Section 2: Hydén's Socio-Legal Concept of Algo Norms

Algo norms as a concept is very new: Hydén coined it in a chapter that was published in December 2020. Hydén's aim in coining such a term is to advance SoL research in the field of new digital technology. For Hydén, researching algo norms allows to demonstrate the normative consequences embedded in this technology (2020, p.363). Thus, the socio-legal research calls for discovering and articulating the advanced practices (ibid, p.364). Nevertheless, Hydén notes that this exercise can start in actual or anticipated actions in order to "articulate [the algorithm's] driving forces, the hidden preferences it generates and their potential consequences" (ibid, p.364; 366). To achieve this, the norm concept has to be extended beyond law and social norms: it "must include technical, economic and professional norms" (ibid, p.367). Put differently, it is key to have a holistic approach when exploring the development of new digital technology, beyond law and social norms; these 'other' norms include the workplace and technical norms.

Section 3: D'Ignazio and Klein's Data Feminist Concept of Strangers in the Dataset

The concept of strangers in the dataset by D'Ignazio and Klein is to demonstrate that developers are usually removed from the data they are working with (2020). The example used to illustrate this was a hackathon in Massachusetts looking at library records in Carolina (ibid, p.132). The librarians in Carolina used "upstate" to define the location of some publications. However, "upstate" is relational to the State; in other words, in Massachusetts "upstate" would have a different connotation, rendering difficult to map out where the publications are across the United States. Only someone within that community would understand "upstate", especially as the information about this is not included anywhere in the library record (ibid). This is to be expected since they are part of that community, although for "strangers", these types of connotations instantly lose their meaning. Hence the term “strangers in the dataset”, since data scientists are not usually part of the community which they are automating systems for (ibid, p.133). Yet, developers are still expected to automate data, which uses terms such as “upstate” by the original community creating this data.

Strangers in the Dataset recognises that developers “must be able to tame the chaos of information overload” (D’Ignazio & Klein, 2020, p.131). This very obligation of their role –to make sense and make use of the data to enable automation— results in developers having to make “deliberate actions” even when they are removed from the data (ibid). This is an issue since they are removed from the original community producing the data and thus might not recognise the “tainted historical roots” of the data (ibid, p.131). This issue is also accentuated if developers use a third party to collect their data, meaning that the developers do not collect their own data but use a dataset from readily available sources online (ibid).

D’Ignazio and Klein warn that being “strangers in the dataset” is ultimately inherently bad (ibid). They draw upon postcolonial scholar Gayatri Spivak’s term “epistemic violence”, to display that developers may reproduce the privileged knowledge (the one where data is collected for) and overlook those oppressed by such systems (ibid).

Section 4: An Operative Framework to Focus on Developers

The theoretical framework appreciates the complexities of developers’ role. Larsson and Hydén demonstrate the need to be wary of norms within AI tools and automated systems, from society itself but also the norms that emanate from the developers themselves. D’Ignazio and Klein look beyond these systems to look at the initial data and the difficulty developers face when trying to use the dataset for automation appropriately.

Merging these theories give an operative framework, which drives this thesis in order to answer the research questions. By appreciating which norms are being mirrored by developers; the algo norms behind the new technologies; and, the intricate roles developers play whilst innovating in old fields, allows to look directly at HRI developers' explorations and involvement in automating fields, such as depression diagnosis in SARs.

These theories provide a general framework. However, since I look at developers in social robotics only, I also allow for some flexibility beyond the framework in order to supplement the theories. This in turn also displays findings from the empirical data which the theories may not account for.

CHAPTER 5: METHODOLOGY

This chapter demonstrates the steps I took in order to answer the research questions. Section 1 looks at the research design; in section 2 I reflect on my own position as the researcher; section 3 sets out my sources of data and how I collect them; section 4 explains the content analysis method I use; and finally, section 5 reflects on ethical considerations.

Section 1: Research Design

The focus of this thesis is on HRI developers, and how their explorations intended to develop social robots further might have normative consequences. This is tied to SARs in the setting of depression diagnosis, to demonstrate normative issues developers will face around embedded gender norms in the diagnosis. Accordingly, my research design has as its object the developers of social robots, which are recognised as experts in their field. What constitutes an expert here is important to mention; in line with Döringer's translation of Kaiser, "experts are considered knowledgeable of a particular subject, and are identified by virtue of their specific knowledge, their community position, or their status" (Döringer, 2021, p.265). Consequently, "expert" is meant for people working on developing social robots, however none of them have much (if any) knowledge on or consideration for medicine and/or gender. This choice was intentional as it was to achieve an explorative epistemological function, enabling to "gain knowledge and orientation in unknown or hardly known fields" (Döringer, 2021, p.266).

In order to collect relevant material for this, I used two different types of research design: ethnographic –by attending the Human-Robot Interaction (HRI) Conference– and conducting semi-structured interviews with social roboticists. The Conference and the experts chosen for the interviews were not necessarily linked (i.e. the expert did not have to attend the Conference), however the knowledge gained from both were interconnected. In other words, the knowledge I gained from going to the Conference enabled me to formulate relevant questions to the interviewees. In the same way, having attended the conference, I understood better the HRI community and the experiences interviewees spoke about.

Section 2: Reflexivity

Usually, reflexivity should come at the end in order to demonstrate how the researcher was reflecting on her position. However, due to my previous experience, my network enabled me to access the HRI field.

I completed an internship alongside Larsson, a socio-legal scholar specialised in AI technologies and its impact on society. Alongside this, I was a project assistant looking at AI in healthcare (on physical health only), participating in projects with computer scientists and epidemiologists. I am now a project assistant for Larsson, alongside a social roboticist professor, Castellano, to create a general framework on gender fairness for social robots, through pilot scenarios.

My current role allows me to get a direct insight into social robots, but it also makes me part of the discourse. I have therefore had to ensure that I separated my two assignments: completing my thesis and the project. I have remained independent throughout my thesis by delimiting my thesis in a critical socio-legal study; whilst my role as a project assistant is to understand how to bridge sociology of law and social robotics to produce a general framework on fairness which focuses on gender. The hope for this project is to be useful at all stages of social robots, in other words, the framework could be used for the development, design, implementation and commercialisation of the social robot in any given context. Whilst my thesis is a critical socio-legal perspective on HRI developers in a very sensitive context, that is depression. This exploration is not within the realm of my work with Larsson and Castellano.

Nevertheless, I asked Castellano to discuss some aspects of social robotics before I began collecting my data. This conversation is briefly mentioned in the presentation of my findings chapter, since it enabled me to understand how to interview/converse with HRI developers; however, it is not part of my dataset. Said differently, I do not use my conversation with Castellano in my findings and analysis. Furthermore, Castellano also acted as a gatekeeper as she gave me contact details for potential interviewees who are developers of humanoid social robots. Getting this type of information and interviewees without her help would have been near impossible. Other than this conversation and gatekeeping, neither Larsson or Castellano were part of the process of this thesis.

Section 3: Source of Data, Collection Methods and Tools

Before I started collecting my empirical data, I organised a video call with Castellano, to gain insight into the role of developers. Castellano was also able to act as a gatekeeper and provided me with contact details of employees at social robotic companies in order to schedule interviews.

My first set of empirical data was collected from a week-long Conference: the HRI Conference. This Conference is a very prestigious event for people interested in social robots: out of 182 papers submitted for the Conference, only 42 papers were chosen to be presented at this year's HRI (ACM Conferences, 2021). Ethnographies conducted at Conferences are not common practice by social scientists, therefore I relied on Supper's way of conducting such an ethnography (Supper, 2012). As Supper points out, this exercise is fruitful since Conferences are where "questions are asked and debates on principles take place, and in which (…) the nature and the boundaries of the discipline are discussed and defined" (ibid, p.30). Accordingly, it is important to record not only the presentation itself, but also the images, gestures and other non-verbal content (ibid, p.31).

Throughout the Conference, I recorded the presentations, the live comments, Q&As as well as my feelings about particular presentations. Although it is important to note that I could not network nor interact with participants in a natural setting, only on the main chat. This meant I could not get any clarifications or fully appreciate the tone of the conversation, which body language usually helps reveal.

After the Conference, I collected my second set of data: expert interviews. I held 3 expert interviews via Zoom which lasted between 60 to 110 minutes, with one 45-minute-long follow-up interview. The participants were all developers and were specifically engineers (or pseudo engineers, ie someone who was doing the tasks of an engineer without the formal qualification). Throughout the interview process, such as contacting the participants, making the interview guide and conducting the interview, I was aware of the sensitivity of depression diagnosis and the embedded gender norms. Accordingly, I based myself on Hove and Anda's study on conducting semi-structured interviews with engineers (2005). In this article, Hove and Anda mention the need to create a safe and trusting atmosphere especially if the participant will be asked sensitive questions. My conversation with Castellano enabled me to understand how to facilitate this setting.

I drafted the interview guide in six parts: getting to know each other; team environment; diagnosing depression; gender norms within depression diagnosis; and closing the interview (see Appendix A). This guide allowed me to have some set questions but also rephrase appropriately. The structure enabled me to first understand how developers view their role to then appreciate how they would approach designing depression diagnosis in a robot. The latter, which I deemed sensitive, included the notion of gender inequality which is what I built my questions on regarding the issue; this approach of keeping the sensitive issue until the end of the interview was advised by Hove and Anda (2005, p.5).

When conducting the actual interview, again in line with Hove and Anda, I ensured that I encouraged the participant to speak freely and phrasing questions in a non-threatening manner nor disagreed with them (ibid). I tried encourage conversations on topics they seemed most interested in and describe those to me before we spoke about depression diagnosis. This was useful as the first part of the interview mapped out the interviewees opinions and values as well as their experiences and thoughts (ibid, p.7). The part on depression diagnosis and social inequality were much more knowledge questions—factual information – since the participants were hesitant to take a stance on such a delicate healthcare matter. Thus, from the interviews, I was still able to understand the potential mirroring of norms and algo norms, and how this could transfer into depression diagnosis and gender questions.

Section 4: Data Analysis

In order to reflect the nuances and the sensitivity of the topic of this thesis, I have chosen to do an ethnographic content analysis (ECA). ECA is part of qualitative content analysis deriving from media and communication. It is an iterative method allowing to demonstrate the nuances of the data and the interaction between the data and the researcher (Altheide, 2016). This means that the context of the data is key and the researcher needs to be aware throughout the procedure (from choosing where to collect data to coding the data) to then enable constant comparison to further delineate specific categories (ibid).

The key to ECA is "the role of the investigator in the construction of the meaning of and in texts" within a given context and social setting (Bryman, 2012, p.291). This enables a systematic and analytic coding but not necessarily objective, since the emphasis is on validity (Altheide, 2016, p.18). In other words, before coding the data I have predetermined some key categories, but once coding is underway, the coding can be amended accordingly – demonstrating the iterative process. To do so, I use NVivo, which is a computer software to code data qualitatively.

ECA allows me to reflect on my own position as a social scientist looking at developers of social robots. There is a need to show my own narrative when I was at the Conference and interviewing, since this will also affect the way I interpret my data. The advantage of this analysis tool is its validity: the circular method of reflecting, interpreting data, as well as reflecting and interpreting the interpretation which ensures the accuracy of my intended measure (Babbie, 2004, p.143). Although validity can never truly be proven (ibid), which is a limitation of ECA.

One of the main issues with ECA is replicability. Since the coding is up to my interpretation, my construct, I need to ensure that how I decide to code can be somewhat replicable. In order to overcome some of this obstacle, I am using a code book with a clear description on my parent and child nodes when coding. My parent nodes make up the umbrella codes such as “source of influence” and “gender”, whilst my child nodes are more specific within a parent node, such as under “source of influence”, my child nodes are “herself”, “community” and, “science-fiction”. Through NVivo I can also create attributes to add context to the interviews and Conferences, for example, which segment the presentation was in (ie HRI Conference session on perception), if the text was from the chat. I am also writing memos to keep track of my line of thinking and reflections. This would enable some replicability and ensures the reliability of my findings (Babbie, 2004, p.142).

Section 5: Ethical Considerations

In line with the Swedish Research Council for Humanities and Social Science, as a researcher I ensured that my study was of high quality whilst also respecting my participants (Swedish Research Council, 2017). I have therefore asked for consent from my interviewees through a written consent form and oral consent at the time of the interview (Appendix B). Once I refined my research question, I asked for consent again; this led to one follow-up interview. I allowed my interviewees’ to choose their name for this thesis and anonymised their workplace.

For the conference, the details of the presentations are available publicly along with the abstracts. Accordingly, any open access information, I do not anonymise since there is not much (if any) ethical considerations. However, for the Q&As and the comments in the chat, I asked for consent from the presenters which would be most relevant to my thesis. I asked three presenters and all agreed for me to use their answers during the Q&A. Any other comment or question interesting to me was directly anonymised by only recording the comment itself.

CHAPTER 6: PRESENTATION OF EMPIRICAL FINDINGS

This chapter represents the core of this thesis. Throughout this chapter, discussions about HRI developers and around gender norms and sensitive issues—focusing mostly on depression—are analysed.

Section 1: Preliminary Discussion Before Starting My Data Collection

Having a social science background, I was unsure how social roboticists work and what their day-to-day tasks might look like. In order to understand this before undertaking my ethnography, I set up a meeting with a Professor at Uppsala University specialising in social robotics, Ginevra Castellano. We agreed to discuss her role as a social roboticist, the high possibility of social robots being programmed to help screen for depression, gender aspects, as well as showing me how roboticists go about experimenting.

The discussion showcased the blurriness of what a social roboticist is, as well as the diversity within the workforce behind making robots:

> Ginevra: we have a team [at the lab] with different kinds of expertise. Most of our students have a computer science and engineering background. And some of them will do more technical work on using machine learning. Others are focused more on setting up human robot interactions to understand how people perceive robots, sometimes we combine the two. But we also have experts from cognitive science and social science in the lab. (…) We have worked a lot with psychologists and education experts [in specific projects]. And then we have different kinds of inputs and different kinds of expertise in the larger setting.

This demonstrates the silos yet dependency between developers. In other words, developers work separately on various aspects of the robots but they depend on the others to ensure the developments function. This resulted in me choosing "developers" instead of "social roboticists" to showcase to the reader the diversity within the workforce.

However, the gender aspect felt much more sensitive. The insight of this conversation made me aware that I needed to be mindful when talking about gender. Firstly, because the notion of gender seemed binary and secondly, gender might be interpreted as the appearance of the robot itself.

Overall, the lab tour enabled me to see how much roboticists enjoy developing software for their robots and coming up with creative fun ideas to develop human skills. Ginevra was very enthusiastic to show the programmes (usually games) that she and her team had created for the robot in order to make the robot more user-centric. Although she made it clear that they enhance one of the robot's capabilities, meaning that they will use the default settings of the robot for the rest. Those default settings will be set up by the companies making the robots.

Section 2: Background Information about My Data

2.1: The HRI Conference

On March 8th 2021, I attended my first day at the HRI Conference— although, no reference to Independent Women's Day was made. This year marked the 16th HRI yearly Conference, which shows the infancy of the community. The theme was "Bolder Human-Robot Interaction" in Boulder (Colorado) – it was undoubtedly a pun, which shows the tone of the HRI community: geeky and fun.

The HRI Conference was over five days. Two days were designated to workshops and three days to plenary sessions. The speakers were usually from academia or research centres. I attended sessions I felt were most suitable to my research, here is an overview:

HRI Conference		
Date	Session	Titles
08/03	Workshop	Child-Robot Interaction & Child's Fundamental Rights
09/03		Blaming the Reluctant Robot: Parallel Blame Judgments for Robots in Moral Dilemmas across U.S. and Japan

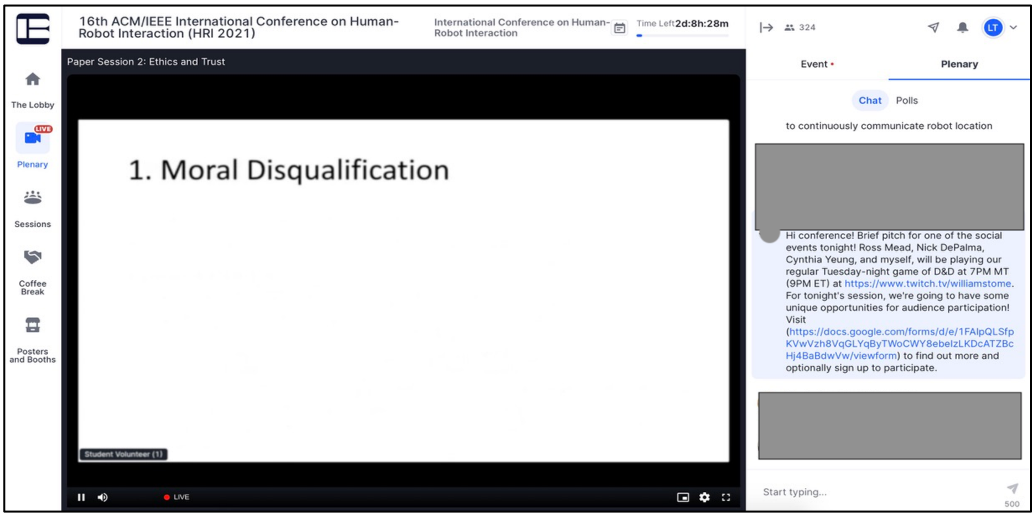
	Plenary: Ethics & Trust	Using Trust To Determine Decision Making & Task Outcome During A Human-Agent Collaborative Task
		Influencing Moral Behavior Through Mere Observation of Robot Work: Video-based Survey on Littering Behavior
		Can You Trust Your Trust Measure?
		Assessing and Addressing Ethical Risk from Anthropomorphism and Deception in Socially Assistive Robots
09/03	Keynote	**Hiroshi Ishiguro**: Constructive Approach for Interactive Robots and the Fundamental Issues
10/03	Plenary: Perception	Uncanny, Sexy, and Threatening Robots: The Online Community's Attitude to and Perceptions of Robots Varying in Humanlikeness and Gender
		Perceptions of Infidelity with Sex Robots in Monogamous Relationships
		"I think you are doing a bad job!": The Effect of Blame Attribution by a Robot in Human-Robot Collaboration
		Effects of Social Factors and Team Dynamics on Adoption of Collaborative Robot Autonomy
		Flailing, Hailing, Prevailing: Perceptions of Multi-Robot Failure Recovery Strategies
		You're Wigging Me Out! Implications of Telepresence Robot Personalizations on Viewer Perception
		What's The Point? Tradeoffs Between Effectiveness and Social Perception When Using Mixed Reality to Enhance Gesturally Limited Robots
10/03	Plenary: Teaching, Learning, & Health	Why We Should Build Robots That Both Teach and Learn
		Effects of Gaze and Arm Motion Kinesics on a Humanoid's Perceived Confidence, Eagerness to Learn, and Attention to the Task in a Teaching Scenario
		The Effects of a Robot's Performance on Human Teachers for Learning from Demonstration Tasks
		Feature Expansive Reward Learning: Rethinking Human Input
		"Is this all you can do? Harder!": The Effects of (Im)Polite Robot Encouragement on Exercise Effort
		Challenges Deploying Robots During a Pandemic: An Effort to Fight Social Isolation Among Children
		Exploring the Design Space of Therapeutic Robot Companions for Children
10/03	Keynote	**Mary-Anne Williams**: Designing Human-Robot Interaction with Social Intelligence
11/03	Plenary: Alt. HRI	Boosting Robot Credibility and Challenging Gender Norms in Responding to Abusive Behaviour: A Case for Feminist Robots
		Who Wants to Grant Robots Rights?
		Robots as Moral Advisors: The Effects of Deontological, Virtue, and Confucian Ethics on Encouraging Honest Behavior
		Sex robots in care: Setting the stage for a discussion on the potential use of sexual robot technologies for persons with disabilities
		Fake It to Make It: Design explorations as research contributions in HRI
11/03	Keynote	**Mark Billinghurst**: Empathic Computing and Human Robot Interaction
12/03	Workshop	The Road to a successful HRI: AI, Trust and ethicS (TRAITS) Workshop

Figure 3: Schedule of Sessions attended at the HRI Conference 2021

From an essentialist perspective, the sessions I attended were represented by both female-presenting and male-presenting persons, with female-presenting persons being more often the speakers. Although it is worth noting that the speakers spoke on behalf of their teams to showcase their current research; thus, there could be a gender dominating in a team and the person speaking was part of the gender minority. However, that did not come across during the conference. Correspondingly, for the sessions I attended, the chairs moderating the sessions tended to be female-presenting persons.

Regarding the keynote speakers, they were academics recognised worldwide within the tech field. Hiroshi Ishiguro is a famous pioneering social roboticist, especially famous for making a replica of himself in the form of a social robot (Guizzo, 2010). Mary-Anne Williams is also a pioneer in the HRI community, who collaborates with entrepreneurs to accelerate innovations in Australia (UNSW Research, 2021). Finally, Mark Billinghurst is outside the HRI community, his research looks into virtual reality, situating itself in HCI (University of South Australia, 2021).

The setup of the conference was well organised. Abstracts for all sessions were available ahead of the presentations, meaning that I could familiarise myself with the content before the actual sessions. The workshops were on Zoom, and the plenary sessions as well as keynote speakers were on a specific platform to be able to access various features whilst watching live presentations (see figure 4). During the plenary sessions, I could pause at any time, enabling me to jot down more thorough notes. If I could not keep up with the presentations and Q&As, I could watch back the presentations later and concentrate on the Q&As as the latter was not recorded. The questions were asked on the chat, so I could pre-emptively organise my notes in order to transcribe word-for-word the speakers' answers.


16th ACM/IEEE International Conference on Human-Robot Interaction (HRI 2021)
International Conference on Human-Robot Interaction
Time Left 2d:8h:28m
324
LT
The Lobby
LIVE
Plenary
Sessions
Coffee Break
Posters and Booths
Paper Session 2: Ethics and Trust
1. Moral Disqualification
Student Volunteer (1)
LIVE
Event •
Plenary
Chat
Polls
Start typing...
500

to continuously communicate robot location

[redacted]

Hi conference! Brief pitch for one of the social events tonight! Ross Mead, Nick DePalma, Cynthia Yeung, and myself, will be playing our regular Tuesday-night game of D&D at 7PM MT (9PM ET) at https://www.twitch.tv/williamstome. For tonight's session, we're going to have some unique opportunities for audience participation! Visit (https://docs.google.com/forms/d/e/1FAIpQLSfpKVwVzh8VqGLYqByTWoCWY8ebelzLKDcATZBcHj4BaBdwVw/viewform) to find out more and optionally sign up to participate.

[redacted]

Figure 4: A Screenshot of the Plenary and Keynote Sessions' Online Setting

The presentations I found most useful for my thesis, and I coded the most on NVivo, were mostly studies based within the EU. Hence the scope of this thesis focusing more specifically on the Western context, especially the EU.

2.2: The Interviews

The second part of my data collection were expert interviews with developers. The interviewees all worked at renowned humanoid social robotics companies. The interviews were scheduled shortly after the HRI Conference to draw comparisons and observations I made, and how this might compare to the industry. The interviews were a lot of fun and insightful, whilst the interviewees themselves were inspirational. Here is a summary of who they are:

Name	Roles	Background	Objective of role
Charlie	Interaction Developer, part of software development.	Philosophy (Diploma—did not complete thesis part), courses in Maths, Computer Engineering and Psychology. Also a famous performing artist using the social robot which he programmes.	Explores possible uses of humanoid social robotics as well as the boundaries of social robots to find interesting learnings.
Karen	Developer in the expressivity team, part of software development.	Physics (Bachelor's), Mechanical Engineering (Master's), Human-Robot Interaction (PhD).	Synthesises multimodal social intelligence for the robots. In other words, making the robot react and interact in a more humanlike way.
Jane	Developer in the agility team, part of software development.	Computer Engineering (Bachelor's), Robotics (Master's), Human-Robot Interaction (PhD).	Understanding how to improve the body movements of the humanoid social robot according to social cues.

Figure 5: Overview of the interviewees

All of the participants were currently based in the EU, although two of them were citizens from outside the EU. Each interviewee was very welcoming and enthusiastic about their area of expertise. They were all very socially aware, in the sense that they were very attentive to me and my own knowledge of the field. If they saw that I did not fully understand some technical points, they would go back and explain before moving onto the point they wanted to make in order to answer my questions. In essence, the developers were very observant especially around body language and non-verbal social cues, despite interviews being conducted on Zoom.

Section 3: The sense of Community

3.1: The Infancy of HRI

> "We're in an industry that does not really exist yet, but there's a high chance in 10 to 15 years, that there will be social robots everywhere" – Charlie.
>
> "As you saw at the HRI [Conference], every year there is lots of research done. You see which one could best fit your solution or to inspire from them."- Jane.

Charlie and Jane's comments demonstrate the way the field of HRI continuously builds from previous research in order to strive and bring social robots into real life applications. This reliance is not competitive, but rather amicable in order to achieve their common goal. To developers, this infancy means that social robots are mostly at research stage, resulting in the aim being more about how to get robots commercialised, rather than how to improve the readily available product. Accordingly, this showcases that their objective is to render social robots legitimate for everyday use; and also, an implicit recognition that developers are *de facto* impacting society since they are creating new inventions. This has been recognised by the engineers at a hackathon (IEEE et al., 2021), and Šabanović (2014).

3.2: A Community Throughout HRI

To achieve their common aim of mainstreaming social robots, there is a clear sense of community and camaraderie. At the opening of the HRI Conference, the Chair stated "instead of enjoying beautiful Boulder at the moment, we are all stuck behind our desks with our coffees, and among our papers instead of among our friends". From this comment, the chat was instantly brought to life with everyone greeting and welcoming each other.

This sense of community also seems to be part of the developers' work culture. For example, there was this sense of co-dependence:

> Tish: (…) because for me, engineers are very much like 'you have your role, but you always come back together'. So you're all interrelated. Is that correct?
>
> Karen: Yeah, we are. So I think in tech everything connects in a way. So you need to think of it as more of a chain and there are different engineers at different parts, but then the whole thing comes together. So we may work individually but we work together at the same time.

At the Conference itself, the speakers shared their code (on an open-source website) at one of the sessions. The justification for this 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. When one member of the audience pointed out that the code was not publicly available, the team rectified it as soon as possible.

The literature did not touch upon the appreciation developers have for another, nor the recognition of everyone's unique skills within the teams. D'Ignazio and Klein for instance draw solely on data analysists as individuals (2020). This might be due to the uniqueness of social robots being a physical artifact as well as software, and a nascent field. However, this respect for one another and camaraderie was very apparent whilst observing the HRI Conference, interviewing and coding my data. Although one of the presentations at HRI specifically warned about this reliance on one another and the need to be more critical of it.

3.3: A Community Entering the Mental Health Setting

Conor McGinn and his team presented "Exploring the Design Space of Therapeutic Robot Companions for Children" at the Teaching, Learning, & Health session. It is worth pointing out that the robot is to assist children and not a humanoid; however the therapy setting shines a light on how developers undertake such a venture and the normative impact of such an invention.

At the beginning of the presentation, McGinn recognises that using SARs to "address symptoms of loneliness, anxiety and social isolation can be especially challenging due to factors that are complex and multifaceted". However, it is important to note that the accent is on the design space itself; thus, I understood the challenges to be around those and not towards the mental issues themselves. Accordingly, McGinn et al's problem statement was "how might robot technology improve the care given to children's hospitals" to be able to make a suitable SAR prototype. They first undertook a literature review, namely on children. Then they conducted observations and interviews with paediatric doctors as well as children specialists at the a children's hospital. Those enabled McGinn et al to map out challenges they needed to be weary of, such as the importance of play but hospitals not having enough resources to provide adequate play to children.

Interestingly, the next step, which preceded the design stage, was "the Ethical Canvas", whereby they landscaped the surrounding of the application:

> The Ethics Canvas is a collaborative brainstorming tool that has the overall aim to foster ethically informed design by improving the engagement of researchers with the ethical impacts on their work.

This Ethical Canvas was not an ethics guideline made by policy makers, but academics. From this Ethical 'Canvasing', McGinn et al drew up a list of what they needed to be aware of and reflect on during the design stage. This included the robot not leading to a reduction in one-on-one time between the medical staff and the patient. Interestingly, no points were critical of the medical institution. Once this was mapped out, the developers took several months to design a SAR, named Taco.

During the Q&A, the Chair was interested to get advice on working alongside medical professionals:

> Chair: Do you have any advice on working with clinicians or folks from the medical field based on the experience developing taco?
>
> Connor: One is to engage with the clinicians at the earliest stage possible. Before you start prototyping or doing any design work, because from our experience, at least in [country], doctors, paediatric doctors in particular, are just so busy. And if you could give them an all-ready design, there's gonna be way too many questions for them to have the bandwidth. So even just having an early conversation to familiarise yourself with them, to get their influence at an early stage, it clears the ground a lot. And also, it's very easy to underestimate the number of stakeholders that can stop you getting into a paediatric setting. So even if you have your consultant paediatric doctors, as we did, there was still all kinds of layers of bureaucracy as well as an oversight that you do need to go through and it can be difficult for a researcher to know that in advance. So my advice would be to engage with them as early as you possibly can. And you can't be surprised with just how supportive they will be but get them on your side, the sooner the better.

This presentation is pivotal to this thesis as it shows that (1) HRI is now entering the medical setting, and McGinn et al are paving the way on how to do such research, (2) developers instantly view themselves as strangers in the dataset in this setting and are merely enablers for this technology, and (3) developers rely and depend on medical staff only in order to develop the appropriate SARs.

There was no mention of critical studies involved, such as gender studies or SoL scholars, in order to reflect on potential obstacles McGinn et al would face. This will make up a part of the algo norm, whereby researchers involve who they feel appropriate for their study. This demonstrates the point made in the literature review: depending on how people view SARs ability in the depression diagnosis setting, they will try and answer an aspect to it according to the problem they set out (Fosch-Villaronga et al, 2020). In the next section, this will be become more apparent along with the normative consequences of this algo norm.

These findings are problematic when comparing it to the literature review on depression. The section looking at depression, demonstrated that medical professionals are reproducing skewed and discriminatory diagnoses based on gender (see chapter 3, section 1 herein). However, McGinn also states that the medical professionals are very busy and have some input so that the prototype McGinn et al will be adequate in the therapy setting. The rest, which makes up the bulk of normative positions, will be made by McGinn et al. Since they are strangers in the dataset, they will likely reproduce the dominating narrative set by the powerful institution, as Benjamin warns (2019). This results in reproducing norms embedded in medical studies, which are likely to oppress people who are not part of the data. Consequently, gender norms discriminatory to men in depression diagnosis will probably be mirrored in SARs.

Section 4: Appreciation of Developers' Own Role

4.1: Process of Developing

In line with Larsson's concept of mirroring norms, and the developers' goal to bring social robots into society, I asked my interviewees what their creations reflected:

> **Overall question asked to all interviewees:** Tish: how you design a robot, is it more reflective of the current social structures, or is it more a reflection of the developers?
>
> Karen: Well, I think the customer and the culture and the human is more important than what the developer thinks. I mean, from the market and from the user, you can understand what's a problem they're facing and how would be a solution for them. But then how you solve that problem is up to the engineers. But defining the problem is up to the user.
>
> Jane: I think it is highly related with the developers. Because you are the first step to how you recognise the things and how you choose the features. But, there were some studies that maybe you also saw in HRI too, where some studies are looking at society [specifically].

> Charlie: I would say that it develops the designers view of society. And there's no way to get out of that. And that kind of bias is very hard to get rid of, even in a robot application as well, and the way to get rid of it is through iteration, and through user testing and talking with people and being very open to the fact that... you don't understand every situation.

These extracts demonstrate the developers' awareness that their creation is part of their own thought process. However, it is clear that the HRI community is keen to embrace this about the developing process and accept the challenge of overcoming it. This is somewhat in line with D'Ignazio and Klein's description of developers: rockstars, ninjas, janitors, unicorns and wizards in order to automate a certain field (2020, p.133); as well as Šabanović calling out researchers making assumptions to advance social robots in society (2014). Nevertheless, to developers this failure to accommodate initially should be celebrated, as they see failure as an inevitability to succeed in time. Put differently, although the part of the robot they enhance is according to their view of society, in time they amend it to fit users better.

This ties with the narrative that developers want to help society generally through social robots. Jane, for example, spends time with the robot to understand its limitations. From there she chooses what is feasible for her to improve, checks the available literature to see if the issue has been solved on other robots. Jane describes this general process:

> That's the experimental process: First you come up with an idea. And then you do something like a "pilot study" like we say. It's something small to see like "is everything going well, as you imagined?" And if everything goes as you imagine, then you go do a bigger study. If it's not, then you tailor it to maybe another small pilot or put it onto the big study.

Clearly "is everything going well, as you imagined?" echoes the mirroring of their own understanding. Consequently, the improvements that they make might be mirroring their own privileges, something D'Ignazio and Klein referred to as a “privilege hazard” (2020, p.29). Said differently, part of the algo norms – or driving force—is their own understanding around how to enhance the robot. However, they are well educated developers who are not representative of society generally.

4.2: Developing and Awareness around Gender

Gender was a very sensitive topic in my interviews. Each interviewee understood gender issues as concerning discrimination of women. Charlie—who has a background in humanities and social sciences— was aware of gender but admitted that sometimes he did not take into account because he was blind spotted by his own gender. Here was his example:

> When we did the unboxing experiment, it was one thing we realised, where one woman that we user-tested on said that very often… like when you place the robot on the table to start it up, it's often at a desk height and you're standing up because there's stuff you need to pick up from boxes. And when the robot is just on the table and it's looking straight ahead— when it wakes up, before it registers anything— it's just kind of looking straight ahead. And she said she felt like it was looking at her breasts. Because that was the height.

In contrast, Karen – a trained engineer– explained to me that it was a shame that I was focusing on gender when there were more important issues to take into account, such as culture, when developing:

> Karen: I don't see gender as something I need to point out to and be like, "oh, my God, I need to design certain gestures specific to women versus men". I really don't. I'm so over this. But I do make sure that I have equal representation of men and women when doing the data collection to make sure there's no bias there. I do make sure that the data I collect is already not biased. If I'm in a context where I need to train data and it's based on human input, I really do try to avoid having data that is very specific to even geo locations. (…) The bigger issue is not gender. The bigger issue is culture.

Later on in the conversation, we discussed more in-depth about gender:

> Karen: But I mean, if you are concerned about gender, make sure women are equally part of the conversation. Make sure they're equally part of the design process. Make sure you have that equal representation on the table and you have to make sure your data isn't biased. And that's the best way you can move forward. For me personally, I think the last thing as women we need is something segregated and be like “This is men. This is woman.”. So just create things for humans. Point. And then actually your humans are represented and that's it.

Jane initially understood my question on gender as related to the way the robot looked. When I asked about gender beyond the aesthetics, Jane explained how gender recognition is used in robots and that it does get the gender wrong sometimes. However, gender recognition seems important because that way the robot is able to recognise an individual more easily. For example, five individuals might look similar and the gender aspect will help differentiate them. Jane was enthusiastic to find out more on the topic of gender and why I was interested in it.

The literature review consisted mostly of either social scientists that were critical of technology or, advancements within the tech field. Accordingly, the former accentuates the gender disparity and gender norms within society which technology amplifies (see chapter 3, section 3 herein); whilst the latter focuses on advancing certain fields and does not pay much attention to gender (see chapter 3, section 2 herein). This subsection shows a bridge between the two fields. The three developers show that they are aware of gender to some extent, and that it needs to be considered, however it does not play a big role. This is an interesting aspect as it seems like they are by default “strangers in the dataset” when looking at gender, since they do not account for it explicitly when programming parts of the robot. Seemingly, even if the developers are part of the community (e.g. unboxing robots) they still have blind spots which affects their programming. This has huge normative effects, which the developers might not realise, especially when accommodating for embedded gender norms rather than explicit gender-related questions.

4.3: Reflecting on Human Interactions

The literature has tended to group developers together. However, it seems that HRI developers are trying to legitimise social robots – making up a big part of the algo norms in this community— which is likely different to developers from other fields. Indeed, the professional norm is to bring social robots into society and thus HRI developers research human-to-human interactions to facilitate robots integrating in everyday mundane tasks. However which human-to-human interaction developers use as a starting point may not be reflective of society. This is probably due to what D'Ignazio and Klein have warned about regarding privilege hazard on behalf of developers, resulting in them not recognising forms of oppression (2020, p.29). This becomes apparent in their process of developing. Here is a sample from Karen explaining an experiment which looks at human interactions, and how her work fits into it:

> Karen: For example, there was one [experiment] on group interaction. So trying to understand how groups are being formed around the robots and according to whom. Because when we are interacting as humans in a group context... So sometimes we're active, but sometimes we're just bystanders to a certain interaction. (…) If you are a bystander, then you [can] jump into the conversation. You [the developers] need to make sure that the robot is also able to adapt in such a conversation. (…) We brought in different social gestures that humans do, different gaze, mechanisms that we do like turn-taking how we subtly say "OK, I'm giving you time to speak", or like, "don't interrupt me, I'm speaking".

Jane furthers this point by showing how developers might try and analyse the data:

> Jane: (…) I conducted an experiment. First, I tried [the experiment myself] and then I asked my colleague to try it and I observe it. Then I asked my colleagues to try it and I observe: "What are the misunderstandable points?". And then after that I see their feedback: "what did you understand from the robot?" Or "what I tried to show with the robot, was it clear or not?". Due to corona, I couldn't go to the real world, but I asked the other departments of the company to do the test of my robot, to collect my data.

As we see from both extracts, trying to understand human interaction is key in order to understand how to develop different facets of the robot; these reflections have an impact on how the developers create and design social robots. HRI developers seem to be aware and reflect on people and how people communicate with one another, in order to try and mirror it into a robot. It is a kind of reflection, in that they view interaction as quite simplistic: one person talks, the others listen by default; or well-educated colleagues might be good enough representatives. However, developers do not reflect on the societal structures which can be biased and skewed, as well as the intricate power plays – such as why is someone talking more and the topic of conversation which might affect the interaction. This lack of societal reflection is touched upon by Benjamin, who warns that overlooking social structures results in reproducing them (2019). This is interesting with regards to embedded social norms: if the developers do not look out for social structures within interactions, they are likely to not recognise oppression against certain genders.

Applying this finding to the theoretical framework, it may be that the developers are somewhat strangers in the dataset through overlooking power relations; particularly as their goal is to clean and use the data to allow the robot to interact. This is part of the algo norms, where developers will create algorithms in line with their findings but might overlook important aspects –such as embedded gender norms. In turn, this will affect the decisions developers make whilst enhancing the robot's capabilities, and will have a normative impact and mirror certain informal norms.

4.4: The Theoretical Implications of Developing SARs in the Setting of Depression

Within the setting of depression, the interviewees immediately admitted that they did not know much about the condition. In this context therefore, they are clearly strangers in the dataset, although they were still willing to demonstrate how they would programme the robot in order to be used in screening for depression. For all three of the interviewees, the robot could pick up on many cues from the patient and build on them to help the doctor with the diagnosis. Karen and Jane drew on the VH experiment mentioned in the literature review, which showed that people were willing to talk to virtual agents (Lucas et al, 2014). This, to them, was a promising start for robots to enter this field. However, to Charlie, this type of development would need to start in research and not from the social robotic companies themselves; so he would not be able to comment on how feasible it would be. Nevertheless, all of them mentioned the importance of having therapists on board to understand how the robot should interact with the user (just like Connor McGinn, no mention of critical studies).

In the follow-up interview with Charlie, we discussed more in-depth social robotic experts and depression tools. Charlie explained that it is not the social robotic expert to question medical professionals and their own expertise:

> Charlie: Currently, if you were making a depression diagnosis application, you would just take an already existing depression diagnosis tool, and basically transport it into a social robot. And probably whatever biases are already in that tool, they would be transferred into the robot version of it. If you were creating a completely new depression diagnosis tool, then the development of that tool would come with its own clinical trials, and all of that stuff that medical professional would be dealing with. And again, that wouldn't really have anything to do with me [as a developer]. The stuff that I would deal with is how does the robot best get the information that the medical researcher wants.

In other words, Charlie explains that it is not the role of SAR developers to question the medical tools in the first place. Instead, developers' role in this instance, is to ensure that SARs are programmed and function according to medical professionals' standards.

All of these responses to mental health settings, including McGinn's presentation as well as the reliance on only medical professionals demonstrates, the lack of reflection on the social structures themselves. According to the interviewees and presenter, the goal is to help the gatekeeper, here the medical professional. However Benjamin points that by attempting to solely take the burden off the gatekeeper will likely to produce an inappropriate model for the new technologies (2019, p.30).

Nevertheless, these findings demonstrate Larsson's observation, whereby developers have "normative positions on issues they would prefer to avoid", which could unintendedly reproduce societal issues (2019, p.590). This shows a direct mirroring of norms, however this mirroring is not of society, it is a mirroring of the privileged institutions in society. This implies that depression in men might not be accounted for in automation in SARs or the way the SARs interacts with an individual, if it is mostly mirroring the current trends on depression; this finding aligns with the two systemic reviews of SARs in a mental health setting, where the participants were mostly women (Abdi et al., 2018; Scoglio et al., 2019). Conversely, this makes up part of the algo norms which influence future developments of SARs. For this to change, social scientists, especially academics from critical studies, would need to be involved during the development process.

Section 5: Developers and (or versus) Guidelines

5.1: An Insight into the Helpfulness of Ethics Guidelines

Whilst I was observing the Conference and interviewing, I wanted to understand if the law, albeit enforceable and non-enforceable regulations, were being followed and helpful to developers.

Katie Winkle et al partly explored this through a specific Ethics Guideline. In the Ethics & Trust session, Winkle presented "Assessing and Addressing Ethical Risk from Anthropomorphism and Deception in Socially Assistive Robots". The aim of this research was to attempt "to navigate the apparent mismatch between typical practice in social human robot interaction, and the first published standard for ethical robot design" (The British Standard BS8611) and seeking a middle ground. The guideline identifies deception and anthropomorphic practices as "ethical hazards" and thus should be avoided.

Using Pepper, which they had already programmed as a fitness coach, they applied the ethics guideline. Through conducting an ethical risk assessment, they found three key risks: deception of the robot such as the robot having feelings; user over-trusting the robot; uncanny valley feelings from the user because of its too-human-like features. To mitigate these risks respectively, the robot should “minimise displays of affect[ion], and to have the robot be very upfront about its real robotic nature”; “the robot [should] refer to appropriate human authorities, and to make sure its capabilities and limitations are made clear to users; and “minimise unnecessary social behaviours and anthropomorphic design keys” (Winkle).

Winkle et al investigated the impact of those mitigations on 120 participants with a robot in various fitness coaching states. In its anthropomorphic state, Pepper would say “I know that exercise can be boring”. Whereas in a lower risk approved by the ethics guideline, Pepper would say “many patients find exercising boring”. They found that users identified deception to be acceptable and that anthropomorphism was somewhat needed. Thus “based on these results, we came to the conclusion that anthropomorphism is important for socially assistive robots, and overall poses relatively low ethical risk, but we do recognize it might alienate some users. Ideally, then, we suggest the display of anthropomorphic behaviours should be tailored to individual user needs and preferences”.

Furthermore, the Chair for the session picked up on Pepper saying that exercise is “boring”:

> Chair: I was wondering if there might be another potential risk there, which is when you say that other people like something, it's kind of establishing a social norm. And whether there might also be some kind of fear that by having the robot refer to them, you might also be biasing people towards a particular kind of attitude or action. In that particular one it might be people might think that exercise, ‘it's okay to think exercise is more boring’, and therefore, give up on it or something like that. So I was just curious if you had thought about the balance of that robot anthropomorphism versus some other kinds of factors that might come into play that you exchange for the anthropomorphism?
>
> Katie: So in this one, we were really just looking at how could we essentially still do the positive social behaviour that a human would do, but in a way that did not require the robot to identify itself as capable of having this type of capability. So

> in this case, it was you know “oh, yeah, I'm not surprised you find it hard because lots of people do”. So we didn't really think about how it could change the norm. But more just how can we still be empathetic without a robot basically being like, “I feel it too”, or like “I understand” when it doesn't. But in their previous work when we did, because we just kind of followed on with work with therapists, they were very good at identifying these risks, actually, in terms of like, “you need to be careful what the robot is saying, because it can normalize”. And actually, it seems, therapists give people a pretty tough time that we wouldn't necessarily think about. So it's definitely something we should consider. But in this case, it was just taking all those classic statements you see on these social robots are like, “yeah, of course, it's hard” and, you know, “I'm really sad to see that you're that you're struggling” and just simply changing that focus to be on the appropriate human or to just take out the arguably deceptive part, but keep the kind of sentiment as best as we could.

Accordingly, we can see here that (1) guidelines are not necessarily mirroring the reality of developing and (2) the ethics guidelines does not necessarily pay attention to social norms. This risk cannot be undermined: what the developer chooses for the robot to say is overlooked by the guideline, even though this could have a normative impact on the field developers are automating— such as exercising. In this setting, it was only reflected upon because Winkle et al worked with therapists. Seemingly, the guideline itself overlooks what the robot may say or how developers could reflect on these.

McGinn et al spoke about their reliance on an Ethics Canvas to design a SAR. During the Q&A, the Chair chose my question to clarify what exactly made up this canvas:

> Tish (read out by the Chair): Did you use any legal frameworks to choose the Ethics Canvas? E.g. the UN Convention on the Rights of the child (which specifically speaks of the right to play and best interest of the child) OR/AND Ethics Guideline such as the one by the AIHLEG set up by the European Commission?
>
> Connor (presenter): Some legal aspect is used to choose the Ethics Canvas. The Ethics Canvas has been something that our research group had come across as very

> useful tool to understand ethics as a process rather than just something you do before an experiment, and reduce each prototype with a little bit. And we thought it was quite useful in this application where there really were many unknowns and it was very hard to predict in advance all of the different ethical factors that would influence things. I actually published a paper on this last year at HRI.

I then checked the paper Connor mentioned and the legal aspect is the actual enforceable laws, such as the GDPR, which partly framed the ethics canvas.

These findings on guidelines are echoed in the interviews and workshops. All experiments developers wished to conduct on humans had to be approved by an ethics board. However, when I asked the interviewees about the standards that they followed during the developing stage, they knew some existed but did not follow any, except the GDPR. Each explained that only when the product is to be commercialised then standards have to be followed – but not at research stage.

It is worth mentioning that Jane did speak about the hinderance of the GDPR by making it increasingly complicated to research in-depth human behaviours outside a lab setting. This leads to unfavourable consequences since Jane cannot collect some vital data for her studies. This may result in collecting data from open sources, which D'Ignazio and Klein warned about due to unknown context and historical roots of the data, leading to further issues with developers being strangers in the dataset (2020); or collecting data from colleagues only, where privilege hazard comes into play as discussed above.

These findings point to the lack of support for developers, which Larsson warns about (2019). Normatively, developers are not guided on how to reflect and make decisions which will not reproduce and amplify societal challenges.

5.2: Developers' View on Policy-Makers

The examples above show the importance of developers' own judgement since the guidelines are not helpful in guiding development— this has already been pointed out by various literature (e.g. IEEE et al, 2021; Fosch-Villaronga et al, 2020). Interestingly, Karen also brings the issue of who is participating in the conversation on AI generally and robots:

> Karen: We're moving more towards decentralized AI. And so 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?". And so we always have this conversation, I'm always invited on panels and I talk and everyone's like: "it's your responsibility to tell the world [about AI]" and I'm like "I can tell the world, but the world is not interested. What do I do? You know how hard it is to get policy-making people on board?". Like suddenly people are interested in AI and I'm like: "hello! Machine learning has existed for over two decades. I'm sorry if it took you like twenty-five years to figure it out, you know. It exists!". So I don't know what you [policy-makers] think we [developers] should do. Hold banners? (very sarcastic tone).

Connecting this section to questions of embedded gender norms demonstrates that the guidelines do not show developers how to challenge current structures. Relating it to Larsson's two questions: "should they [developers] reproduce the world in its current state or as we would prefer the world to be? And who gets to decide which future is more desirable?" (2019, p.590); it appears that for now, HRI developers have to make these normative decisions alone. Firstly, because the guidelines might not take into account the typical practices – as shown by Winkle. And secondly, because the policy-makers and others outside of development, are not keeping up with the technological advancements. Consequently, developers have to make normative decisions, which are unlikely to challenge the status quo as they are not trained in this. Accordingly, Hydén's algo norms does have to go beyond the law, since it will not necessarily drive the development and creations.

Section 6: The HRI Conference Challenging and Guiding Developers

6.1: Alt.HRI Challenging the Status Quo around Gender

One presentation was specific to questioning the role of gender in HRI at the Alt.HRI session, called "Boosting Robot Credibility and Challenging Gender Norms in Responding to Abusive Behaviour: A Case for Feminist Robots", presented by Katie Winkle. The research was based on UNESCO's report "I'd blush if I could" which exposes the lack of gender diversity among the tech industry and developments (UNESCO, 2019). Winkle et al transferred these findings to robots: "[i]n this work, we set out to investigate whether we could improve perception and effectiveness of such a robot by actively going against these norms". Accordingly, Winkle defines a feminist robot as "any robotics activities that name and challenge sexism and seek to create more just equitable and liveable futures" – inspired by *Data Feminism* (D'Ignazio & Klein, 2020).

Winkle et al had three aims stated below:

> Winkle: Firstly, we use a robot to explicitly encourage girls to consider studying robotics and have the robot express a feminist sentiment in this context. Secondly, we consider if and how a robot should respond to negative anti-feminist sentiment and direct insults or abuse in this context. Finally, we utilise a female stylised robot to deliver aggressive and argumentative responses to this abuse, specifically going against subservient female persona, typically occurrent issue assistance, as well as human cultural norms regarding female politeness.

There were three randomised groups of high school students, where the robot (with female-looking attributes) will respond differently to each group. Winkle et al relied on the University's outreach/advertising material which attempts to bring more women into STEM subjects. Accordingly, the robot repeated the following: less than 30% of people working at [name of University] are women, as well as the slogan "the future is too important to be left to men". After the robot said these, a male actor will shout abusive or sexist comments – these responses were drafted by teachers. Depending on the experimental group, the robot would respond one of three ways to the abuse: “I won’t respond to that” (control group); argue back; or become aggressive.

Winkle et al found that the robot could challenge sexist biases. Girls found feminist robots more credible, and for boys this did not impact their perception of the robot. However, “we didn’t get everything right” Winkle states. They found, for example, that in the aggressive condition, girls became significantly disinterested in robots. One reason, they suggest, is that “by focusing on girls in the aggressive condition actually highlights the risk of further marginalisation”.

During the Q&A, someone asked whether it was appropriate to rely on gender issues regarding the lack of women in social robotics. Winkle answers that it raises questions on how outreach is done, “particularly around trying to encourage young women and people who are non-binary into doing this”. To try and bypass this issue next time, the team will use a participatory design to overcome this obstacle. Winkle’s closing remark was a reflection that herself and the team appear to have solely concentrated on the appearance of the robot. However, they are aware that a feminist robot is beyond the way the robot looks, but it was a “good starting point”.

This presentation uncovers how developers will look at and transfer currently used material into robots. This is a facet of strangers in the dataset, whereby developers rely on existing material to build a part of the robot. Winkle shows awareness of the downfalls of this, whilst also pointing out the added process to rectify the real-world material outside of robotics. This circles back to Larsson’s concept of mirroring norms, whereby developers reproduce and amplify current practices prior to automation. Here we see a need for more critical studies to be involved in the development process to avoid discriminatory norms—especially key in the setting of depression diagnosis.

The audience was ecstatic about this presentation and thanked them for having brought up such a challenging topic. This shows a need to include more critical studies in HRI. As we can see from my interviews, gender is something the HRI field is just starting to look into, with Winkle paving the way. My findings have also pointed to HRI community learning from one another, which is an algo norm that veers HRI developments. Accordingly, there is a possibility that this presentation will be reflected in other developer's future work.

6.2: Alt.HRI Guiding Mindful Development

Although it has been touched upon that the community is welcoming of other disciplines to learn how to advance social robots—especially HCI— the HRI community is also willing to learn beyond engineering. A great example of this was seen at the Alt.HRI, which looked specifically at a sensitive topic where there is a huge need to reflect on what should be developed and how to avoid reproducing biases and stigma. The subject here was sex robots for people with disabilities. Eduard Fosch-Villaronga (Associate Professor in Law) and Adam Poulsen (PhD candidate in computer science focusing in value sensitive robot design and LGBTIQ+ eldercare), presented "Sex robots in care: Setting the stage for a discussion on the potential use of sexual robot technologies for persons with disabilities".

Fosch-Villaronga first centred around sex and people with disabilities outside of technology, to be critical of development in this area. Fosch-Villaronga reflected on the current state of society: "[a]lthough sexuality is a central aspect of human experience, awareness and knowledge do not come straightforward for disabled populations". He points out that this may be due to "largely constrained pathologized and ignored in different health settings". Also, social and structural factors will also add burdens to people experiencing mental illness, such as stigma and the medication. Nevertheless, sexual needs are universal, and this raises two questions about sexual needs and sex workers accord to Fosch-Villaronga:

> Firstly, it is not clear the necessary knowledge required for satisfying the sexual needs of a disabled person and whether there should be a minimum safeguard.

> Second, sex work is controversial concerning sex workers legal status and the lack of educational licence structure.

Sex robots can directly contribute to the "improvement in the satisfaction of essential needs of a user" as well as being directly adapted to the needs of the person. However, Fosch-Villaronga notes that an issue is that sex robots are not usually targeted at people with disabilities but "young and non-disabled and typically straight men". Nevertheless, the functions of the sex robot could still be fulfilling and will not require much physical activity from the person if required.

However, sex robots could also reinforce societal biases relating, for example, to sexism or machismo. The cost of sex robots may also contribute to further socio-economic disparity with only a certain group accessing the sex robot. There is also the aspect of "humanisation of caring practices" which is essential in the care setting. Though, care homes have often overlooked sexual needs of their patients. Thus, the inclusion of sex robots could instead "help realise the recognition of the sexual rights of the users and make their practices more humane". To achieve this, developers need to be mindful of these structures when programming. The starting point, according to Fosch-Villaronga, should not be in engineering solutions, but start by reflecting and consider society to accommodate the diversity between humans to critically anticipate the design and implementation of sex robots in care.

Fosch-Villaronga's presentation can be applied to the depression setting, whereby developers need to be aware of stigma and biases despite the lack of literature on the subject on how to programme this into SARs. This does render developing intricate and complex, and as Jane pointed out in the interview, all of a sudden engineers have a similar status to doctors whereby they hold people's lives in their hand, yet, unlike doctors, they do not know how to deal with these processes.

Fosch-Villaronga's training shows how to accommodate for issues where developers are strangers in the dataset. Fosch-Villaronga offers a critical insight around

Larsson's reflection that developers will have to take a normative stance: if done mindfully, the design process can reflect the needs of the users in vulnerable and taboo situations.

This section demonstrates that developers' role is overlooked, including by guidelines, when they are wanting to know how to achieve a mindful and useful robot design. The HRI community is welcoming to everyone, including those without much programming experience. This is an algo norm, whereby they are willing to invite people who will help find solutions outside of technology to current societal issues. This shows a growing awareness to the fact that developers may need to take normative stances. And as we have seen, this is key when they may replicate the status quo in a depression setting. Thus it is hugely important to support developers in their role, especially as development continues in the setting of depression.

CHAPTER 7: CONCLUSION

This thesis' main aim was to bring forth HRI developers' role in creating and programming robots, which may be used to screen for depression. Although there were some limitations to this study (section 1), section 2 answers the research questions, to then offer policy recommendations fit for developers in section 3.

Section 1: Limitations of the Study

The methodology used for this thesis was hybrid: conducting an ethnography and interviews. This led to a breadth of data, which were not always related to one another. This was partly accommodated for by the ECA coding method and delimiting the setting to the EU context mostly. However, as the researcher, I had to choose which part of my data to present and how to present it. Accordingly, my findings are tentative answers to my main research question, and more studies on this topic would need to be undertaken to be able to generalise my findings.

Another limitation is due to the scope of this thesis, whereby I had a vested interest in exploring socio-legal issues in a STEM subject. Thus, I applied my knowledge and experience to the subject of SARs and the setting of depression. Although I tried to be reflexive on this, my biases inevitably veered my research.

The final limitation concerns SARs in the depression setting: they are not yet commercialised and are still at research phase. Consequently, HRI developers do not tend to closely follow regulations – this cannot be generalised to technology that is widely commercialised. More research would need to be undertaken as well as interviews with specific HRI individuals who work on commercialising SARs.

Section 2: Answering the Research Questions

The research questions centred around the developers to then focus on the exploration of SARs in the setting of depression diagnosis. There were two sub-questions in order to answer the overarching question.

2.1: Sub-question 1 – Developers Advancing HRI

The first sub-question was embedded in the theoretical framework and looked at how developers were advancing the HRI field. The findings emphasise that developers from the HRI community collaborate and cooperate with one another. To them, it is the norm to look at what the others are doing to apply it to their own project; this is viewed as advancing HRI by developers. Accordingly, regardless if they are strangers in the dataset, they will look for inspiration—especially within the HRI community—to pick the best solution to their exploration. This mutual assistance has two-fold normative consequences. Firstly, the developers find something within the social robot which they want to enhance; this first step means that they are reflecting themselves into the creation – something that they are aware of. Secondly, they look at possible solutions that have already been found, usually because developers are strangers in the dataset and want inspiration to get around the data they have – this is well demonstrated by handing out codes at the Conference. This second step shows a mirroring norm but not that of society as Larsson states, but the mirroring norms of the other developers' understanding of society. This understanding rooted from the other developers will be used by the exploring developers, however it will also be adapted to the developers' own understanding of the dataset and society's need. This reliance and adaptation is a driving force within the community, which forms an important part of the algo norms.

2.2: Sub-question 2 – Regulations to Develop and Design SARs

The second sub-question was embedded in SoL generally, which transpires through Larsson's mirroring norms (2019). Part of SoL's aim is to recognise informal norms which exist alongside legal norms and are dependent on the context in which people find themselves in. Larsson points out that these informal norms have to be accounted for, since they will be embedded in datasets. Although Larsson mentions those with regards to data, it can also be applied directly to developers. Since legal norms, recognised as regulations in this thesis, tend to be value driven, especially in the area of new digital technology they are not useful to developers whilst exploring (see chapter 2, section 3 herein). As a results, the HRI community is mostly following informal norms, since the legal norms are not reflective of their practices and what they are trying to achieve with the robots. This is an interesting finding in itself, since the regulations are seemingly more hindering than helping developers. Consequently, the HRI community relies on itself to decide how to develop and design social robots (and by default SARs).

2.3: Main Question – The Role of Developers themselves, focusing on the gender norms embedded in Depression Diagnosis

The core of this thesis is to bring forth developers' role from their perspective and showcase the consequences of it in the setting of depression diagnosis. Depression diagnosis, as demonstrated in the literature review, has seemingly embedded gender norms which hinder the diagnosis of men with depression. This hinderance is part of the medical institution which tends to emphasize research on women in this setting. In turn, this causes the dataset to have "tainted historical roots" (D'Ignazio & Klein, 2020, p.131). However, developers tend to view their roles as enablers, meaning that developing SARs in this field does not involve critically evaluating the context of medical institutions. Instead, the role of developers —according to them— is to ensure that the robot can assist the medical staff to make their job easier. This very narrative has been picked up by Benjamin, whereby technical solutions are said to take the burden off gatekeepers, but overlook social divisions which the new technology consequently reproduces (Benjamin, 2019, p.30).

These findings insinuate that gender norms embedded within datasets are overlooked. This was well illustrated by Winkle's presentation, on making people aware of gender issues, and the interviewees reactions to gender questions. Developers do not feel equipped to tackle these questions – they are strangers in the dataset. Thus, although they reflect on the goal that they want to achieve as HRI developers, they do not reflect on societal structures and the normative decisions they are making. This merged to the depression diagnosis setting amounts to developers feeling ill-prepared to tackle these issues around embedded discriminatory norms. As a result, the findings identify that HRI developers generally do not consider their role as influencing the design of SARs in depression settings. This is in part due to the algo norms and the mirroring of norms, as discussed above. However, in reality developers have to make important normative decisions which they will mirror within the SARs.

Section 3: Policy Recommendations

This thesis has pointed to the lack of support for developers, whilst in parallel the expectations on developers by regulations. These are well documented through various guidelines who expect developers to collect the correct data or arrange the data in such a way that is in line with specific values. Developers are continuously expected to be ahead of society in order to create new useable technologies. However, these guidelines have been called out by developers as not being helpful during the experimental process. As we saw with Conor McGinn and his team, in order to use SARs in therapy settings, they relied on an Ethics Canvas to oversee potential obstacles – and not guidelines from a regulatory body. Regardless of the usefulness of guidelines, developers still have to make normative decisions which have normative consequences. Furthermore, none of the interviewees were aware of guidelines to help them develop. It seems that policy-makers are overlooking the needs of developers, for now. Accordingly, I propose seven recommendations:

1. **Regulations should accommodate beyond AI tools**

A holistic approach which accommodates for the way data is collected, how developers are expected to use data and recognise the vital exploration phase. The current EU draft on AI regulation mentions code of conducts for developers generally – this is a promising start (European Commission, 2021). Additionally, there is a need to appreciate that social robots only use some facets of AI and will also use other types of software to function.

2. **Define developers in regulations**

Developers are not a one-size-fits-all. HRI developers in the industry and HRI developers in research centres differ slightly from one another despite being in the same field. One concentrates on commercialising a default social robot, whilst the other wants to enhance a particular characteristic of the robot. This also changes with regards to humanoid HRI developers and other types of robot HRI developers.

3. **Educate developers in critical studies**

HRI developers are ill-equipped with regards to questioning societal structures, especially in a trusted institution such as medicine. The HRI community recognises this and invites people from all disciplines to accommodate this shortfall; however, it should be supported by society generally in order to reflect on which norms should be mirrored in social robots.

4. **Bring policy-makers into the conversation**

Policy-makers need to be present in conversations with developers first-handed to ensure that guidelines, recommendations and laws reflect the reality of developing.

5. **Recognise the limitations of current developers**

D'Ignazio and Klein referred to data scientists as rockstars, ninjas, unicorns, janitors and wizards (2020, p.133). Although developers are very skilled and inspiring, they should not bear the burden alone to accommodate for society. In turn, social scientists – such as SoL and gender scholars – should be trained to understand new digital technologies so that they can help tackle these increasingly important questions around biased societal structures. Hopefully this will allow social scientists to be part of the development process.

6. **Raise the expectations to challenge the status quo**

As it stands, regulations allow for developers to reproduce the current societal structures. Even the obligations from the GDPR results in developers having to collect their data from third parties in order to continue their experiment. This has normative consequences and does not allow developers to critically reflect on the data they have collected. This can have harmful results, as demonstrated in the setting of depression diagnosis and the embedded gender issues.

7. **Concrete formulations around the process of exploring and challenging the status quo**

This thesis demonstrated that developers have to create and design ahead of knowing what society might need in the future. However, they are usually automating raw data from the real-world. Ideally, there should be some guidelines to demonstrate developers how to question the current norms from the setting they are automating. This is key, as it is at the point of exploring that many norms are established and reproduced by developers.

Section 4: Concluding thoughts

My title, *Behind every social robot finds itself a community of developers: A socio-legal exploration on developers of humanoid social robots with a focus on the context of depression diagnosis and its embedded gender norms*, reflects the journey I embarked on to answer my curiosity around developers and mental health. Before this thesis, HRI developers seemed unattainable, with knowledge I would never be able to acquire. Although the latter part may be true, HRI developers have been incredibly welcoming and open-minded about my own research. They were candid about their own shortfalls, which included questions of gender and embedded discriminatory social structures; to me, this reflection shows the potential willingness to critically assess the automation of certain institutions. Throughout my research, it has become clear that it takes a community to bring robots into real-life application. Now is an opportunity to expand that community.

REFERENCE LIST

Abdi, J., Al-Hindawi, A., Ng, T., & Vizcaychipi, M. P. (2018). Scoping review on the use of socially assistive robot technology in elderly care. *BMJ Open*, *8*(2), 18815. https://doi.org/10.1136/bmjopen-2017-018815

ACM Conferences. (2021). *HRI'21: Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction* . Association for Computing Machinery. https://dl.acm.org/doi/proceedings/10.1145/3434073

Altheide, D. L. (2016). Ethnographic Content Analysis. In *Qualitative Media Analysis* (pp. 14–23). SAGE Publications. https://doi.org/10.4135/9781452270043.n2

APA. (2020, October). *What Is Depression?* American Psychiatric Association. https://www.psychiatry.org/patients-families/depression/what-is-depression

Atwood, N. A. (2001). Gender bias in families and its clinical implications for women. In *Social work* (Vol. 46, Issue 1, pp. 23–36). http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L33437839

Babbie, E. (2004). Conceptualization, Operationalization, and Measurement. In *The Practice of Social Research* (10th ed., pp. 118–150). Thomson Wasworth.

Bacigalupe, A., & Martín, U. (2021). Gender inequalities in depression/anxiety and the consumption of psychotropic drugs: Are we medicalising women's mental health? *Scandinavian Journal of Public Health*, *49*(3), 317–324. https://doi.org/10.1177/1403494820944736

Ballantyne, A. J., & Rogers, W. A. (2011). Sex bias in studies selected for clinical guidelines. *Journal of Women's Health*, *20*(9), 1297–1306. https://doi.org/10.1089/jwh.2010.2604

Bartneck, C., & Forlizzi, J. (2004). A design-centred framework for social human-robot interaction. *Robot and Human Interactive Communication*, 591–594.

Benjamin, R. (2019). *Race After Technology: Abolitionist Tools for the New Jim Code*. Polity.

Brommelhoff, J. A., Conway, K., Merikangas, K., & Levy, B. R. (2004). Higher

Rates of Depression in Women: Role of Gender Bias within the Family. *Womens Health*, *13*(1), 69–76.

Bryman, A. (2012). Content analysis. In *Social Research Methods* (4th ed., pp. 289–308). Oxford University Press.

Capek, K. (2004). *R.U.R. (Rossum's Universal Robots)* (C. Novack-Jones (Ed.)). Penguin Classics. https://www.penguinrandomhouse.com/books/286379/rur-rossums-universal-robots-by-karel-capek/

Carey, M., Jones, K., Meadows, G., Sanson-Fisher, R., D'Este, C., Inder, K., Yoong, S. L., & Russell, G. (2014). Accuracy of general practitioner unassisted detection of depression. *Australian and New Zealand Journal of Psychiatry*, *48*(6), 571–578. https://doi.org/10.1177/0004867413520047

Criado Perez, C. (2020). *Invisible Women: Data Bias in a World Designed for Men by Caroline Criado Pérez*. Penguin Random House.

D'Ignazio, C., & Klein, L. F. (2020). *Data Feminism*. The MIT Press.

Dautenhahn, K., & Billard, A. (1999). Bringing up robots or-the psychology of socially intelligent robots: From theory to implementation. *Third Annual Conference on Autonomous Agents*, 366–367.

Döringer, S. (2021). ' The problem-centred expert interview '. Combining qualitative interviewing approaches for investigating implicit expert knowledge. *International Journal of Social Research Methodology*, *24*(3), 265–278. https://doi.org/10.1080/13645579.2020.1766777

Dublin City University. (2021). *Johann Issartel | Researcher Details | Dublin City University* |. https://www.dcu.ie/researchsupport/research-profile?PERSON_ID=1631127

European Commission. (2020). *White Paper On Artificial Intelligence—A European Approach to Excellence and Trust: Vol. COM(2020)*. https://www.cambridge.org/core/product/identifier/CBO9781107415324A009/type/book_part

European Union. Regulation Of The European Parliament And Of The Council Laying Down Harmonised Rules On Artificial Intelligence (Artificial Intelligence Act) And Amending Certain Union Legislative Acts, (2021).

European Union. General Data Protection Regulation (GDPR) , Pub. L. No. Regulation (EU) 2016/679, OJL (2016). https://gdpr-info.eu/

Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and Autonomous Systems*, *42*(3–4), 143–166. https://doi.org/10.1016/S0921-8890(02)00372-X

Fosch-Villaronga, E., & Albo-Canals, J. (2019). "I'll take care of you," said the robot. *Paladyn, Journal of Behavioral Robotics*, *10*(1), 77–93. https://doi.org/10.1515/pjbr-2019-0006

Fosch-Villaronga, E., Lutz, C., & Tamò-Larrieux, A. (2020). Gathering Expert Opinions for Social Robots' Ethical, Legal, and Societal Concerns: Findings from Four International Workshops. *International Journal of Social Robotics*, *12*(2), 441–458. https://doi.org/10.1007/s12369-019-00605-z

Furhat Robotics. (2021). *Homepage*. https://furhatrobotics.com

Girgus, J. S., & Yang, K. (2015). Gender and depression. *Current Opinion in Psychology*, *4*, 53–60. https://doi.org/10.1016/j.copsyc.2015.01.019

Guizzo, E. (2010, April 23). *Hiroshi Ishiguro: The Man Who Made a Copy of Himself - IEEE Spectrum*. IEEE Spectrum. https://spectrum.ieee.org/robotics/humanoids/hiroshi-ishiguro-the-man-who-made-a-copy-of-himself

Healthline Editorial Team. (2018, September 16). Depression Tests and Diagnosis. *Healthline*. https://www.healthline.com/health/depression/tests-diagnosis

High-Level Expert Group on Artificial Intelligence. (2019). *Ethics Guidelines for Trustworthy AI.*

Hove, S. E., & Anda, B. (2005). Experiences from conducting semi-structured interviews in empirical software engineering research. *Proceedings - International Software Metrics Symposium*, *2005*(October 2005), 10–23. https://doi.org/10.1109/METRICS.2005.24

Hydén, H. (2020). Sociology of digital law and artificial intelligence. In J. Přibáň (Ed.), *Research Handbook on the Sociology of Law* (Issue 2018, pp. 357–369). Elgar. https://doi.org/10.4337/9781789905182.00037

IEEE. (2021). *IEEE at a Glance*. IEEE.

IEEE, ANE, & University of Copenhagen. (2021). *Addressing Ethical Dilemmas in AI : Listening to Engineers*. 1–38. https://nordicengineers.org/wp-content/uploads/2021/01/ethical-dilemmas-in-ai-xxx-report-final-single-pages.pdf

IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. (2019). *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*. https://standards.ieee.org/content/ieee-standards/en/industry-connections/ec/ autonomous-systems.html

Islam, M. R., Kabir, M. A., Ahmed, A., Kamal, A. R. M., Wang, H., & Ulhaq, A. (2018). Depression detection from social network data using machine learning techniques. *Health Information Science and Systems*, *6*(1). https://doi.org/10.1007/s13755-018-0046-0

ISO. (2021). *ISO - Standards*. International Organisation for Standardisation. https://www.iso.org/standards.html

Keller, E. F. (1987). The Gender/Science System: or, Is Sex To Gender As Nature Is To Science? *Hypatia*, *2*(3), 37–49. https://doi.org/10.1111/j.1527-2001.1987.tb01340.x

Larsson, S. (2019). The Socio-Legal Relevance of Artificial Intelligence. *Droit et Société*, *103*(3), 573–593.

Lucas, G. M., Gratch, J., King, A., & Morency, L. P. (2014). It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior*, *37*, 94–100. https://doi.org/10.1016/j.chb.2014.04.043

Mitchell, A. J., Vaze, A., & Rao, S. (2009). Clinical diagnosis of depression in primary care: a meta-analysis. *The Lancet*, *374*(9690), 609–619. https://doi.org/10.1016/S0140-6736(09)60879-5

Mokhtar, T. (2019). Designing Social Robots at Scales Beyond the Humanoid. In *Social Robots: Technological, Societal and Ethical Aspects of Human-Robot Interaction* (Springer, pp. 13–35).

Moravec, H. P. (2021, February 4). *Robot | Definition, History, Uses, Types, & Facts* |. Britannica. https://www.britannica.com/technology/robot-technology

NHS. (2019a, December 10). *Diagnosis - Clinical depression* . National Health Service . https://www.nhs.uk/mental-health/conditions/clinical-depression/diagnosis/

NHS. (2019b, December 10). *Overview - Clinical depression*. National Health Service. https://www.nhs.uk/mental-health/conditions/clinical-depression/overview/

NHS. (2019c, December 10). *Symptoms - Clinical depression* . National Health Service. https://www.nhs.uk/mental-health/conditions/clinical-depression/symptoms/

Oliffe, J. L., Rossnagel, E., Seidler, Z. E., Kealy, D., Ogrodniczuk, J. S., & Rice, S. M. (2019). Men's Depression and Suicide. *Current Psychiatry Reports*, *21*(10), 1–6. https://doi.org/10.1007/s11920-019-1088-y

Purtova, N. (2018). The law of everything. Broad concept of personal data and future of EU data protection law. *Law, Innovation and Technology*, *10*(1), 40–81. https://doi.org/10.1080/17579961.2018.1452176

Reece, A. G., & Danforth, C. M. (2017). Instagram photos reveal predictive markers of depression. *EPJ Data Science*, *6*(15). https://doi.org/10.1140/epjds/s13688-017-0110-z

Rippon, G. (2019). *The Gendered Brain: The new neuroscience that shatters the myth of the female brain*. Vintage Publishing. https://www.amazon.se/Gendered-Brain-neuroscience-shatters-female/dp/1847924751/ref=asc_df_1847924751/?tag=shpngadsglede-21&linkCode=df0&hvadid=476487651513&hvpos=&hvnetw=g&hvrand=7206187522983989187&hvpone=&hvptwo=&hvqmt=&hvdev=c&hvdvcmdl=&hvlocint=&hvlocphy=1012442&hvtargid=pla-699945496007&psc=1

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

Scoglio, A. A. J., Reilly, E. D., Gorman, J. A., & Drebing, C. E. (2019). Use of social robots in mental health and well-being research: Systematic review.

Journal of Medical Internet Research, *21*(7), 1–14. https://doi.org/10.2196/13322

SoftBank Robotics. (2021). *SoftBank Robotics - Group | Global Site*. https://www.softbankrobotics.com/

Supper, A. (2012). *Lobbying for the Ear: The Public Fascination with and Academic Legitimacy of the Sonification of Scientific Data*. Universitaire Pers Maastricht.

Swedish Research Council. (2017). *Good Research Practice*. https://www.vr.se/download/18.5639980c162791bbfe697882/1555334908942/Good-Research-Practice_VR_2017.pdf

Topol, E. (2019). *Deep medicine: How artificial intelligence can make healthcare human again*. Basic Books.

Trinity College Dublin. (2021). *Conor Mc Ginn : Department of Mechanical, Manufacturing & Biomedical Engineering - Trinity College Dublin*. https://www.tcd.ie/mecheng/staff/mcginnco/

UNESCO. (2019). I'd blush if I could. In *UNESCO - EQUALS Skills Coalition* (Vol. 306). https://unesdoc.unesco.org/ark:/48223/pf0000367416.page=1

University of South Australia. (2021). *Mark Billinghurst*. University of South Australia. https://people.unisa.edu.au/Mark.Billinghurst

UNSW Research. (2021). *Mary-Anne Williams*. UNSW Research. https://research.unsw.edu.au/people/professor-mary-anne-williams

Walther, A., Grub, J., Ehlert, U., Wehrli, S., Rice, S., Seidler, Z. E., & Debelak, R. (2021). Male depression risk, psychological distress, and psychotherapy uptake: Validation of the German version of the male depression risk scale. *Journal of Affective Disorders Reports*, *4*(January). https://doi.org/10.1016/j.jadr.2021.100107

Woebot. (2021). *Mental Health Chatbot | Woebot Health*. https://woebothealth.com/

World Health Organisation. (2020). *Fact sheet: Depression*. https://www.who.int/news-room/fact-sheets/detail/depression

APPENDIXES

Appendix A: Interview Guide

Part 1: get to know each other

Goal:

- Put them at ease
- Present myself
- Present them
  - Education
  - Why social robots
  - Why [social robotics company]

Part 2: Team environment

Goal:

- Describing work environment
- Who is part of the team?
- What have they worked on?
- Who takes responsibility for what?
- What standards/laws do they follow?

Questions:

- What does your role involve?
- Who makes up your team?
  - How many other engineering teams are there at social robotics company]?
  - As a company, what would you say social robotics company]'s end goal is?
- As an individual, what would you say your personal end goal is as a developer?
- When developing/altering aspects on [social robot], how do you decide what needs to be done?
- What standards/law do you follow?
- What aspects of [social robot] been reconfigured? If many, which do you find most interesting?

Part 3: Diagnosing depression

Goal:

- Involvement in mental health
- Current projects?
- [social robot] role in diagnosing depression.

Questions:

Thinking specifically about diagnosing depression

- How would you define depression?
- Do you think it is a good use of social robots to help diagnose depression?
- Are you involved in projects to enable [social robot] to diagnose depression?
- What do you consider in the design process?
- What standards/law do you follow?
- How do you decide on the data that needs to be collected?
- how do you get data?
- What data do you get?
- How do you/would you decide on your measures do find specific depression diagnostic?

Part 4: Social Inequality with a focus on gender within depression diagnosis

Goal:

- Reflection on conversation, how much was gender mentioned?
- How is gendered taken into account?
- As a developer, do they feel a role towards it?

Questions:

This is dependent on the conversation… but!

- Looking back at our conversation so far, how much consideration do you give to gender when programming [social robot]?
- Do you think gender plays a role in depression diagnosis? *(I'm aware this is a yes/no, but I want the individual to reflect)*
- How highly do you rank gender when considering what needs to be accounted for in programming [social robot]?

- Would you say your dataset and how [social robot] interacts reflects society? OR would you say your dataset reflects a segment of society and the developers' views on the matter?
- Is there a possibility that this revolutionary social robot, in your opinion, embed continuities of gender bias?
- As a developer and robotic researcher, do you think it's your responsibility to reflect on gender norms? Whose responsibility should it be?

Part 5: closing the interview

Goal:

- Thank interviewee
- Can send over a transcript within the next two weeks.
- I will not publish the interview in its entirety, only some parts if needed in my thesis.
- I will do a text analysis on the interviews I conduct and the HRI conference

Appendix B: Interview Consent Form

Research title *(subject to change)*: Reflecting what? A socio-legal perspective on the developers' role in creating humanoid social robots in the context of pre-existing gender norms embedded in depression diagnosis research.
Research purpose: Master's thesis in Sociology of Law.

Researcher: Laetitia Tanqueray
Research participant: **X**

Description of thesis: This thesis primarily aims to bring forth the developers of social robot's view and understanding of their current role in society as well as their own experiences. In order to do so, the researcher conducts interviews with developers and attended the HRI Conference 2021. The socio-legal relevance finds itself at the lack of formal legal recognition of developers' role despite developers having important normative powers.

This thesis will attempt to demonstrate this by focusing directly on the possibility of advancing social robots in the realm of medicine, by enabling social robots to help diagnose depression. The reason for choosing this is to demonstrate the societal inequality, specifically gender inequality, in depression diagnosis already existing in medicine. This thesis hopes to demonstrate the underappreciated responsibility put on developers and the potential reflection of current practices reproduced in social robots by developers unintentionally.

Consent to take part and use the interview transcript

I, X, volunteer to participate in a research project conducted by Laetitia Tanqueray, based at Lund University, Sweden. I understand that the project is designed to gather information about my experience as a developer of social robots. I will be one of around three people being interviewed for this research.

1. The interview is recorded and a transcript will be produced.
2. I will be sent the interview transcript and have the opportunity to correct and clarify within one weeks of being sent the transcript.
3. I have had the purpose and nature of the study explained to me in writing and I have had the opportunity to ask questions about the study during the interview.
4. I understand that I can withdraw permission to use data from my interview after the interview, in which case the material will be deleted.
5. The actual recording of the interview will be deleted upon completion of the thesis.
6. I understand that the researcher will identify the necessary information about my role as a developer.**
7. My participation in this project is voluntary. I understand that I will not be paid for my participation. I may withdraw and discontinue participation at

any time without penalty. If I withdraw from the study, it will be stated in the thesis.

8. I have read and understand the explanation provided to me. I have had all my questions answered to my satisfaction, and I voluntarily agree to participate in this study.
9. I have been given a copy of this consent form.

** A few details about the role of the developer needs to be explained to show why the interviewee reacts/says certain things. This is how you will be described in the thesis (please decide the name that you would like):

___________________ is an interaction developer at a leading social robotic manufacturing company as well as a performing artist. ___________________ nearly completed a Bachelor's in Philosophy (he is missing the final thesis) and took courses in Maths and Computer Science as well as other various subjects. As an interaction developer, his skills resemble those of a software engineer, although he does not formally refer to himself as such. His role in the company is very experimental in that he explores possible uses of social robotics as well as the boundaries of social robots to find interesting learnings.

OR

______________ is a research software developer at a leading social robotics company, working specifically on a humanoid social robot. ____________ holds a bachelor's in computer engineering, a master's in Robotics and finishing her industrial PhD in Human-Robot Interaction at a prestigious University and at the social robotics company. More specifically, __________ works in the agility team within the software development department, whereby she looks at how to improve the body movements of the humanoid social robot according to social cues.

OR

_____________ is a research software developer at a leading social robotics company, working specifically on a humanoid social robot. ___________ holds a bachelor's in physics, a master's in mechanical engineering and finishing her industrial PhD in Human-Robot Interaction at a prestigious University and at the social robotics company. More specifically, ___________ works in the expressivity team within the software development department looking at synthesizing multimodal social intelligence for the robots. In other words, making the robot react and interact in a more humanlike way.

Please note: this description is elaborate; it will either be this long or (most likely due to word count restriction) be shortened.

________________________ ________________

Name of participant Date Signature

_________________________ __________________

Researcher Date Signature

To be counter-signed and dated electronically.