Gender Fairness in Social Robotics

Exploring a Future Care of Peripartum Depression

Tanqueray, Laetitia; Paulsson, Tobiaz; Zhong, Mengyu; Larsson, Stefan; Castellano, Ginevra

Published in:
Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction

DOI:
10.1109/HRI53351.2022.9889468

2022

Document Version:
Peer reviewed version (aka post-print)

Gender Fairness in Social Robotics: Exploring a Future Care of Peripartum Depression

Laetitia Tanqueray
Department of Technology and Society
Lund University
Lund, Sweden
laetitia.tanqueray@lth.lu.se

Tobiaz Paulsson
Department of Information Technology
Uppsala University
Uppsala, Sweden
tobiaz.paulsson.2376@student.uu.se

Mengyu Zhong
Department of Information Technology
Uppsala University
Uppsala, Sweden
mengyu.zhong@it.uu.se

Stefan Larsson
Department of Technology and Society
Lund University
Lund, Sweden
stefan.larsson@lth.lu.se

Ginevra Castellano
Department of Information Technology
Uppsala University
Uppsala, Sweden
ginevra.castellano@it.uu.se

***Abstract*—In this paper we investigate the possibility of socially assistive robots (SARs) supporting diagnostic screening for peripartum depression (PPD) within the next five years. Through a HRI/socio-legal collaboration, we explore the gender norms within PPD in Sweden, to inform a gender-sensitive approach to designing SARs in such a setting, as well as governance implications. This is achieved through conducting expert interviews and qualitatively analysing the data. Based on the results, we conclude that a gender-sensitive approach is a necessity in relation to the design and governance of SARs for PPD screening.**

***Index Terms*—gender norms, socially assistive robots, peripartum depression, governance, gender fairness, socio-legal robotics**

I. Introduction

Peripartum depression (PPD) affects around 10% to 20% of women anytime from the time of conception to one year after giving birth [1], [2] and manifests in the pregnant woman with feelings such as of loss of interest and enjoyment [2]. It is a serious condition with high societal costs which can also lead to preterm delivery, adverse birth outcomes, low quality of maternal life, family breakdown, and even increased risk of suicide [3]. In order to receive treatment, a PPD clinical diagnosis is required. This currently involves a structured clinical interview with a skilled physician. However, access to skilled personnel in primary care can vary substantially, which can lead to long waiting times or low diagnostic accuracy. Recent research suggests that up to 69% of PPD cases go undetected and only 6% receive adequate treatment [4].

In this paper we investigate the use of socially assistive robots (SARs) to support medical professionals in the screening of PPD and explore how a future of care of PPD could look like in the next five years from a governance of robots and artificial intelligence (AI)'s perspective. SARs in mental healthcare have, of recent, started to receive a lot of attention [5], whether as tools to monitor and improve people's mental health during infectious disease outbreaks [6], or as positive psychology coaches to support students' wellbeing [7]. However, no previous work, to our knowledge, has investigated the use of SARs in the screening and diagnosis of PPD.

PPD is a sensitive application area for SARs. Pregnancy is filled with expectations, due in part to the *gender norms* on the pregnant person. In this paper we will refer to pregnant persons and new mothers as "women" and use the pronouns "she/her". Within this terminology we include non-binary people as well as trans men and cis-women. In this context, gender norms are not a binary divide between "man" and "woman" but about the social meaning ascribed to a pregnant body and how this can affect the norms around it [8]. Gender norms can manifest themselves through people saying to the pregnant woman "being pregnant is such a gift". However, women experiencing PPD might not reciprocate these feelings, yet may react according to societal norms [9]. This may lead to pregnant women experiencing PPD to suppress their feelings which might go unnoticed by their close entourage, including medical professionals.

Building on previous research that highlights the importance of involving users and key stakeholders in co-designing SARs [10] and on recent ethics guidelines to develop ethical and trustworthy AI systems that promote non-discrimination and fairness [11], [12], in this paper we advocate for a gender-sensitive approach to the design of SARs for PPD screening. Based on the definition of fairness in AI guidelines [11], [12], we conceptualise *gender fairness* as the principle according to which SARS are designed following a gender-sensitive approach. Through our interdisciplinary research team, made-up of social roboticists and socio-legal scholars, we identify how to study gender norms in a PPD setting and how this can be translated into the design of SARs according to the principle of gender fairness which accounts for gender norms.

To inform the design of SARs for PPD screening, we conducted an interview study with experts from two different

fields: PPD experts – people who work in research or practice within PPD – and gender studies experts – researchers in critical studies with expertise on gender. This allows a critical take on PPD and its screening. The aim is to explore how human-robot interaction (HRI) can (1) challenge current medical practices, and (2) be designed to avoid transferring and mirroring norms which hinder women opening up about their current mental health. We discuss HRI design implications for SARs in PPD screening, including contexts of use within and outside of the medical institution, the different roles that SARs could play, and which physical appearance and capabilities SARs should have. We reflect on these implications using a gender fairness and socio-legal lens, and discuss how SARs can impact the future of governance of PPD screening and AI and robots from an ethical and trustworthy AI perspective.

II. Related work

A. PPD and the Swedish Context

Major depressive disorder (MDD) is reported as being mostly experienced by women [13]. In accordance with this, the literature, as well as policy recommendations, are based on women and the symptoms women experience [14]–[17]. Consequently, the narrative around depression is already very gendered. Interestingly, how much MDD is attributable to PPD is not fully known [1]. Yet, literature shines a light on pregnant women feeling stigmatised and isolated about their depressive symptoms [9]. Thus pregnancy seemingly adds an intricate layer which is mostly focused on pregnant women and yet also prevents them talking about their mental health openly.

In Sweden, pregnancy is closely monitored by an assigned midwife [18]. After birth, the woman will usually see the midwife once more, and then be assigned a Child Health Services Nurse [18]. Currently in Swedish national recommendations, women should get screened for PPD eight weeks after they have given birth through Edinburgh Postnatal Depression Scale (EPDS) questionnaire [19]. If the woman feels that her mental health is being affected by the pregnancy, she has to go and see her General Practitioner (GP) to get screened for PPD through what is called the MINI interview or the EPDS; from there she can then get the medication or mental health support. However, in Sweden it is reported that immigrant people living in Sweden have poorer access to healthcare, including mental healthcare [18].

B. SARs in Mental Healthcare

SARs have, of recent, been rising in prominence in mental healthcare [6]. Using SARs has been shown to be beneficial in many ways, such as addressing labour shortage, reducing biases in the diagnostic processes from healthcare personnel, and improving self-disclosure by enhancing a feeling of anonymity [5], [20]–[22].

Other research utilised SARs to support care and management of dementia and autism, and assessment of cognitive impairments [23]–[26], to provide social support and companionship to reduce loneliness [27], or to assist in the education of children with developmental disabilities [28].

Virtual agents have been used in depression assessment. Some applications involved the automatic assessment of mental disorders, including depression, by means of semi-structured interviews [29], [30]. An evaluation showed that users preferred to communicate with virtual agents compared to human interviewers [29]. These studies show the potential of an intelligent automatic diagnostic system of depression.

However, no previous work has investigated robot-assisted diagnostics or support of perinatal depression. Moreover, effective ways to design robots as tools for mental healthcare and to integrate them into current medical practice still require exploration [21], [31], and we draw from our earlier work in this study [32].

C. Social Robotics and Gender

Designing SARs according to a gender-sensitive approach is becoming increasingly relevant [33]. Yet, the study of robot gendering is still underexplored [34], and the question of whether and how a robot should be gendered, in its appearance and behaviour, still open.

Should robots be gendered in the same way that humans are gendered? Should they mirror human gender norms? Researchers are challenging the approach of a binary gender representation of robots, suggesting that robotic embodiments offer the opportunity for a more fluid approach to robot gendering [33]. A key challenge here is that gendering robots mirroring human norms may lead to the reproduction and perpetuation of gender biases. A recent UNESCO report demonstrated that the gendering of voice assistants led to production and reproduction of gender stereotypes, especially the notion that women should be submissive, polite and patient [35]. In social robotics, studies have explored the relationship between robot gendering and gender and occupational stereotypes [36], and perception of robotic noncompliance [37]. Moreover, recent studies investigated how social robots could be designed to go against current digital assistants' gender norms and advocate for a feminist approach to the design of social HRI to challenge and reduce gender biases [38].

We posit that addressing the question of whether and how to gender robots cannot be done without a cross-fertilization of perspectives from different research fields. Indeed, it has been shown before that a dialogue between experts in HRI and gender studies can lead to a more gender-sensitive approach to the design of social HRI [38]. In this paper, we adopt a socio-legal HRI perspective to the design of social robots in the context of PPD.

D. Governance

Sociology of Law (SoL) allows for a wider understanding of governance [39], taking into account regulations and guidelines, but also the norms around how these types of governance are viewed and put into practice. While governance of AI and robotics is a wide and multifaceted field, some say its underdeveloped in relation to recent technical developments [40], for example in relation to robotics and autonomous systems in social care [41]. Our study touches on both a) established

regulations, primarily related to privacy, the care context and anti-discrimination (including gender fairness), as well as b) notions of AI governance, such as the *Ethics Guidelines for Trustworthy AI* published by the EU Commission's High-Level Expert Group on AI (AI HLEG) in 2019 [11], as well as c) the European Commission's proposal for an AI Act.

Firstly, and for Sweden and the screening of PPD, the National Board of Health and Welfare (Socialstyrelsen) has since 2010 recommended that the self-assessment scale EPDS should be included in the conversation with new mothers about their well-being [19]. In addition, a key regulatory framework for privacy and data protection is of course the European General Data Protection Regulation, GDPR, ultimately echoing the right to private life as a human right and aspects of patient autonomy [42].

Secondly, the introduction of AI-systems and machine learning capabilities in health care and medicine points to the relevance of the AI HLEG's guidelines, framing "trustworthy AI" as found in the respect for human autonomy, prevention of harm, fairness, and explicability – including questions of explainability and accountability [**?**], [43]. The ethics guidelines present seven requirements for "the realisation of trustworthy AI": 1) Human agency and oversight; 2) Technical Robustness and safety; 3) Privacy and data governance; 4) Transparency; 5) Diversity, non-discrimination and fairness; 6) Societal and environmental well-being; 7) Accountability. These guidelines are part of a wider trend on defining principles and guidelines [44], [45] linked to the development and implementation of AI and machine learning in various fields, including social robotics, and health [46].

Thirdly, for a European context this has had a particular focus on trust and human-centricity [39], [47], demonstrated by the proposal for an AI Act, published in April 2021. The proposal is roughly divided into AI-systems bearing unacceptable risk, high-risk – likely to include much public sector health care – and low-risk. Commentaries sees the risk-approach as appropriate and practical but consider the wide definition of AI-systems as highly problematic [48], describing the proposed act as "AI auditing" [49], with a great deal of self-enforcement. The proposal includes distinct demands on dataset quality (for high-risk applications), on representativeness, accuracy, completeness, and on human oversight [50].

E. Social Structures, Power and Intersectionality

Critical scholars shine a light on the false promise around the narrative that integrating technology will optimise and render the process more efficient. However, it can also oversimplify and amplify social problems [39], [51]–[53]. This is central to this paper, since pregnant women are often prevented to speak about their negative feelings because they do not want to appear as a "bad mom" nor do they completely understand their own condition [9]. Therefore, to produce an appropriate model, the focus should be on understanding those structures and potential social divisions, before reducing them to algorithms [51], [53].

The role of power, such as the stakeholders involved, should also be accounted for. Pinpointing power enables to account for where the data is coming from (e.g. from only medical data) and whether it is inclusive and representative (e.g. do the patients feel that they are listened to and recorded adequately by their medical professionals) [52]. Put differently, it is key to identify if the stakeholders involved for the development of new technologies are representative and inclusive. For this research, the interviews explicitly included questions which demanded a reflection on social structures and power dynamics within the medical institution (see results and discussion).

Understanding the complexity of direct robot users is vital; as each pregnant woman has her own life-experience and unique background. This is coined as "intersectionality" [54], whereby the individual's unique traits – such as her ethnicity, her socio-economic status– as well as societal factors, will impact women's experiences of the medical institution. The concept of intersectionality demonstrates that a generalised approach to PPD is not appropriate, and thus also veered the research design of this paper.

III. Method

A. Research questions

We investigate the following four research questions:

- Q1: How are gender norms at play in the context of PPD?
- Q2: How can current medical practices in PPD be challenged by socially assistive robots?
- Q3: How can gender fairness be integrated in the design of socially assistive robots for PPD screeening?
- Q4: What aspects of governance are of the most relevance for ensuring gender fairness in socially assistive robots?

B. Participants

Eight participants overall (F:7, M:1, average age: 47.75 $\pm$ 18.75): four with expertise in PPD referred to as "P" in the results (N=4, F:3, M:1; 10-17 years of experience, mean: 14.5) and four with expertise in gender studies from academia referred to as "G" in the results (N=4, F:4, M:0; 3-35 years of experience, mean: 21). All were recruited via email in Sweden.

C. Procedure

Semi-structured interviews were conducted through Zoom by the first and second author in English (N=6) and Swedish (N=2). Prior to the interview study, participants filled in a questionnaire, including demographic information, questions on previous experience with robots and digital technologies, and professional experience.

In the first part of the interview, participants were asked questions on gender norms in current medical practices surrounding screening of PPD. Afterwards, in order to set the context for the next interview stage, using a similar approach to Serholt et al. [55], we asked participants to read a fictional vignette. The vignette was a made-up scenario about a pregnant woman, Molly, experiencing PPD and going to her GP to discuss about it. With her consent, she spoke to a SAR about her feelings and was informed that her GP would review the

discussion. The appointment lasted over an hour and enabled Molly to open up. The full text of the vignette is available in the supplementary material.

Participants were then asked to watch a video demonstration (also available in the supplementary material) where a fictional patient interacts with a Furhat robot (https://furhatrobotics.com/) in an interview scenario for PPD screening (Figure 1). The animations of the robot were implemented with FaceCore Face Engine (https://docs.furhat.io/facecore/), which can capture real human facial animations, voice, and head movements and reproduce them on a Furhat robot. The voice was processed to be more synthetic-like. Following this, the second part of the interview investigated experts' views on how to challenge existing norms and current medical practices with SARs, and what capabilities the latter should have. Participants were then asked to watch a second video, which showed multiple combinations of voices and appearances of the Furhat robot. Specifically, we demonstrated the Furhat robot with a) voice that was congruent and not congruent with its appearance, in terms of gender; b) gendered and gender-fluid appearances; c) human-like and nonhuman-like appearances. Finally, in the third stage of the interview, we asked questions on what characteristics in appearance a SAR should have when used to support screening of PPD.

Each participant received a gift voucher as compensation for their participation. The research was approved by the local ethics committee.

Fig. 1. Fictional patient interacting with a Furhat robot.

D. Analysis

Applied thematic coding (ATC) was used to code the data [56]. It is primarily an inductive, deriving from social sciences as a qualitative analysis tool, and draws on established theme-based techniques, namely grounded theory, thematic analysis and phenomenology [56]. This coding method enabled to bridge the researchers' disciplines as well as ensuring rigour and replicability. The codes were split into structural, based on the actual interview guide, to see the "link between data collection and the evidence generated" [56, p. 77]; and content, based on the context [56, p. 49], which results in demonstrating the link "between evidence and its significance" [56, p. 77], to achieve themes and compare between participants [56, ch. 3]. The coding process was conducted by the first and second author, both of which come from separate disciplines, the former from SoL and the latter from HRI.

IV. Results

A. Negative Gender Norms within PPD

1) Emanating from the Medical Institution: The role of the medical institution is key throughout the pregnancy in Sweden. Thus, the pregnant woman will go to these institutions physically, which one gender expert points out, leads to the person seeing posters of visibly pregnant women with a "hopeful smile". Furthermore, the pregnant woman will only get the option to be screened for PPD through the EPDS questionnaire 8 weeks after giving birth, but not all women fill out the questionnaire and neither is everyone offered the screening [19]. This indicates the priority given to the physical health over patient's mental health, as pointed out by a few PPD expert participants. One PPD expert participant mentioned the lack of time to ask patients about their mental health during the visits; whilst another spoke about their patients having bad experiences with the healthcare personnel. A gender expert, in a similar line, pointed that setting up a GP appointment is already time consuming (where a MINI interview could be set up), which adds to the negative experience.

2) Emanating from the Pregnant Woman: Participants were very sympathetic to the pregnant women. One PPD expert explained that the pregnant woman often blamed herself for developing PPD, whilst in contrast if she developed a physical condition during the pregnancy, she understood and accepted that that was out of her control. Gender experts referred to the gender norms emanating from the pregnant woman as her not knowing how to process or acknowledge her own feelings around PPD. This narrative was echoed by the majority of PPD experts and furthered. For example, two PPD experts explained that the change from the midwife to the nurse checking on the baby's development after birth meant that women wanted to present themselves as good mothers; or when filling out the EPDS questionnaire she would not answer accurately because she did not feel comfortable doing so or she thought that she would get better soon.

3) Emanating from Society and Societal Expectations: All participants put emphasis on the role society plays around PPD. There are expectations from many sources, which point to the importance of being a "good mom" and presenting as such. One aspect derives from media outlets, such as films and social media, where the pregnant woman on the screen seems to be happy and *"having the best time of [their] life"* (P3). Another aspect which leads onto this, is that the pregnant woman does not relate to such content but feels that she has to act in line with it to fit in. A final important aspect, which only gender experts touched upon and some PPD experts implicitly referred to, is this notion that there is one type of desirable mother. This means, in Sweden, that a mother should ethnically, culturally, and from an age perspective be aligned

with the values of the country. If someone falls outside of that scope, they will likely face difficulties. For example, a trans man will not be recognised by the medical system as being able to be pregnant.

B. Challenging PPD Processes with SARs?

1) PPD experts view SARs as an Ally to Medical Professionals: PPD experts were typically welcoming of the robot, especially in the PPD setting. These experts were typically critical of the screening tools used in Sweden for PPD. All PPD experts agreed that the tools (EPDS and MINI interviews) are not tailored to pregnancy specifically. For example, the MINI questions, although they will still be successful in identifying people with PPD, are related to weight and sleep – both of which are deemed to change during the pregnancy and after giving birth. The EPDS does not differentiate between how the mother felt and how the mother felt if the child is not considered with regards to loss of enjoyment. From a governance perspective, the experts pointed to the fact that the medical institution and staff do not have the adequate resources for PPD. Very often it is for the professional to raise awareness around PPD and ensure that the pregnant woman feels ready to speak about how she feels– which takes time and resources. In line with this, experts felt that the robot could challenge these norms and fill the gap and be an ally to the staff.

The PPD experts saw the robot as an integration of their work within their medical team. The findings point towards the PPD experts being willing to bring robots into the PPD setting. This is partly due to the fact that the GP would still be in the loop and would still be the one making the final call, and that the robot has more time to allow the patient to feel comfortable and raise awareness around PPD and is also compassionate and empathetic, which we discuss in more detail in Section IV-C. PPD experts also enjoyed the aspect of the robot being able to go outside of the usual medical procedure, such as ending the session with a breathwork exercise. Reservations on the robot had more to do with how to personalise the robot to the user, as discussed in Section IV-C. One PPD expert also voiced concern over the success of the robot; during the pandemic, her clients were not willing to talk to her virtually. Thus, the pregnant woman, may want the physical assurance from a human in such a circumstance, and not be offered to talk to a robot.

2) Gender experts view SARs as a Reproduction of the Medical Institution: On the other hand, gender experts were much more critical of introducing the robot in the PPD setting. A good summary of the gender experts is the following extract, said in an ironic tone: *"Sure, you could use it [the robot] to challenge norms if you want to, or you could use it to reproduce the existing norms and practices that are already there. I mean, that's one of the freedoms with robots, right?"* (G3). From a sociological perspective, gender experts spoke about the problem around calling PPD "depression". This label has a lot of stigma attached to it, which derives from the medical institution. Put differently, only the medical professionals in Sweden have the power to label somebody with depression and the connotations around such a condition is stigmatised. Thus creating a robot in such a setting is already reproducing the mere act of the patient having to turn to the medical institution for some help. Furthermore, the medical institution has tended to only screen pregnant women and not the partner or other individuals who might suffer from PPD. Thus to them, this research was problematic in part due to this.

From a governance perspective, the role of a primary practitioner is required in order to diagnose PPD. To the gender experts, the robot might merely add a layer of waiting time in-between calling for an appointment regarding the mental health of the pregnant woman and the GP reviewing the transcripts. As a result, this might merely extend the time frame until the pregnant woman will receive help, which does not challenge PPD as it currently stands.

A majority of participants mentioned that the women want to talk to somebody about how they are feeling but they do not know how to address it. They pointed to the robot having similar characteristics to that of something people confess to, or look to in order to openly speak about their feelings. Thus the robot could be developed in a broader context, where the stakes are not as high nor as stigmatised, and which can provide additional support to the pregnant woman. Accordingly, instead, some gender experts suggested developing robots to support people during and after their pregnancy, outside of the medical institution altogether. One gender expert mentioned that in Sweden there are other structures to provide support for the mother and try to find women who need more support. These institutions are more helpful to the medical institution in this instance, in her opinion, since the woman might be struggling mentally not because of depression but other factors, such as requiring additional resources due to unstable living conditions or an abusive partner.

C. Informing HRI Design

1) Overall view of the robot in a healthcare setting: After showing the two demos of the robot, we asked each participant their opinion on them. The first demo regarded the interaction between the robot and user, here participants felt able to comment as experts. However, the second video concerned the appearance of the robot and participants felt that pregnant women should be the ones prioritised.

A majority of the participants agreed that it is key to ask the patient consent about whether she wanted to speak to the robot. Only PPD experts were supportive of the robots in the demos, although they were still reserved: *"Okay, it's a quite new thought to me this robot thing. But yeah, I think [it went] better than expected."* (P3). However, all participants were openly critical about the Furhat robot in the first video demo, for example the robot seemed too happy, or too monotonous, or too robotic and synthetic.

2) Robot's behaviour: PPD experts tended to reflect on the verbal aspect of the robot, whilst gender experts on non-verbal aspects. Gender experts pointed a lot to the need for the robot to show empathy and mirror the user to facilitate this, such

as mirroring user's emotions and behaviours, or nodding. One went further and explained verbal acknowledgement to what the user was saying, such as *"I'm sorry to hear that" or "Yes, I hear that you're saying that this is really bad"* (G3). Most gender experts also pointed to the impossibility of a value-free robot, where cultural context will be key and thus not necessarily generalisable outside of Sweden and European customs around childbirth impacting gender norms.

PPD experts on the other hand were more focused on the verbal aspect of the robot's behaviour. Seemingly, the PPD experts rectified or validated the way the robot behaved towards the user according to how human experts should behave, for example: *"not automatically mirror the patient's behaviour, but (...) check if you have understood what the patient has said. So I think that's pretty important."* (P1). The consequence here is two-fold: the robot should mirror the patient, but also speak with a certain intonation and filler words such as "uhm" (agree) so that the user feels listened to. One participant pointed to the need for the developer to ensure that the robot constantly re-adapts to the patient, whereby the patient might be sad one moment, and the next be happier.

3) Robot's appearance: Generally, the participants were unsure about what the appearance of the robot should be in such a setting, for example: *"My most personal spontaneous feeling would be to be quite absurd to sit and talk to that little plastic head"* (G4). Thus, for the robot to seem credible to the users, it would need to be personalised for such a setting. All felt that in this setting the appearance of the robot would play an important role, but would be very subjective to the person, and could include factors related to both ethnicity (i.e., skin colour) and gender. Usually, participants felt that the robot needed feminine traits, since (a) midwifes in Sweden are usually cis-women and (b) people tend to view women as better listeners and more caring concluding that: *"from experience, unfortunately, I think people would prefer a woman if you have to choose a gender"*. However, this did not mean that the robot should have features of a woman's face, nor the need for strict congruency between the voice and the face. For example, some viewed the man with a beard and a male voice as a credible robot in this setting as it reminded them of good childhood experiences; whilst others viewed that exact robot replicating patriarchal structures: *"the harshness of the voice more dominating (...) [it] reminded me of older generation doctors"* (P2). This disparity seemed to emanate from their own experience.

Furthermore, reflections on anthropomorphism in the robot's appearance were also interesting. According to the participants the robot should have human-like features, but only up to a certain extent. In other words, the robot should resemble somewhat a human, but in a synthetic way, and not have an anime/manga face. Participants agreed that the robot could also be gender neutral in appearance, and have a gender-neutral name and a non-binary voice.

Importantly, it was hard for participants to pinpoint the user's agency in deciding the robot's appearance. Whilst ideally the user would choose how the robot could look like, the very issue with PPD is loss of interest and difficulty to make choices. Thus, participants usually concluded that either the user does not get to choose the pre-set face that is used or the user chooses from a handful of clear choices. But this decision is very value-driven around what the developers think would fit such a setting.

D. Robot Governance and Gender Fairness

From a governance perspective, questions around the learning data is not fully addressed by regulations and guidelines, nor executing gender fairness. Usually governance guidelines surround issues of privacy and the personal data used in training a potential machine learning-based tool. Accordingly, we asked participants what they thought the robot should have access to for PPD screening purposes and why they believed it was important.

As previously mentioned, the PPD experts welcomed the robot into their healthcare practice, thus they viewed the robot as a colleague and as such should access what they can access. To PPD experts, it seemed key that the robot needs to have basic information about the patient, such as their relationship status, their feelings about the relationship and also understanding how to adapt to the patient such as the way they speak to adapt to it, and the intonation and speed of the speech, maybe even language. Whether the robot should have access to the medical records directly was dependent on their professional experience: *"personally when I meet the pregnant woman, I do not have access, I do not even want to read someone else's note, because I want to form an idea of how the patient is doing. So for that reason, I think the robot does not automatically need access to other people's notes."* (P1); whereas in contrast: *"if they're [the robot] part of the medical team, they should get access to the records. And they should know that this is in strict confidentiality."* (P2). Another PPD expert voiced that the patient should consent upon meeting the robot whether the robot should have access to the medical records. However, one PPD expert explained the issue around the decentralisation of medical records in Sweden. Meaning that it is very difficult to access medical records from another region in Sweden, and impossible to access past medical records from outside of Sweden. Thus, to another PPD expert, the robot should directly ask about any hereditary of psychiatric diagnosis. Usually all agreed that the robot should remember past meetings to adapt to the patient, and some experts even saw it as more important than having access to medical records. For "other", one PPD expert thinks that the robot could train on actors to learn how to adapt to different kinds of patients, with different types of background. Another said the patient should fill out a form on an electronic tablet about some information on themselves which would be useful for the robot to know – something another PPD expert felt the robot could ask directly.

Gender experts were usually welcoming of the robot being configured to ask the patient questions to know what values the robot should emphasise on to make the user feel more comfortable – especially if they do not have a good experience

with medical staff. However, one was very critical of such a practice since it tries to reduce the patient to an individual; this does not account for the other people involved in her pregnancy, such as her partner/family/friend who support or hinder her health. Consequently, *"the patient becomes an extended, multi bodied patient."* (G2). Whereas with questions around what data should the robot have prior to the meeting with the user, the gender experts were very sceptical and brought up ethical questions. One issue surrounded the issue of reducing bodies to numbers and the developer having the power to decide what each value means and its importance. Two gender experts saw medical records as being biased to the medical staff writing the record, thus it can be too messy and riddled with different meanings to be useful. Two pointed to privacy issues. Three gender experts either explicitly stated or insinuated that the robot should have access to some data about the patient that is relevant as well as remember the patient, so that it can act in line with the ideal of a medical professional.

V. Discussion

A. HRI Design Implications

1) Contexts of Use and Roles of SARs in PPD screening: This study set out to understand how robots could be included in the screening stage for PPD and help support medical professionals as well as patients in Sweden. We found much more than this. Firstly, the results pin-point that the context of the robot is key due to problematic gender norms emanating from two main areas: society through certain expectations on the pregnant woman and also how she should act, as well as the power the medical institution holds, resulting in pregnant women experiencing negative gender norms from medical institutions. These findings are in line with the notion of "bad mom" and not fully understanding how they are feeling [9]. This result raises the question: should SARs be incorporated in such a sensitive setting? The results from gender study experts would point more towards a negative answer, in part due to their macro-view on this. To these experts, robots might add a layer of hassle to the pregnant woman and reproduce the medical institution merely by being placed within it. Seemingly, placing a SAR into such a setting would reproduce the power the medical institution holds, and in turn the robot might propagate its power and further social division [51], [52] as well as reduce individuals to an oversimplification of what they are experiencing [53]. However, developing SARs to be used during and after pregnancy to support mothers outside of the medical institution is worthy of investigation. Gender studies experts were also positive around the robot still being an entity patients could openly speak to about what they are experiencing. Especially if it is programmed to account for the intersectional aspect of each individual [54]: the gender the individual identifies with, her ethnicity, her socio-economic background, her relationships and many other. Whilst in contrast, to PPD experts, the SAR in such a setting could be useful to them as medical professionals and to the users. The robot could be integrated to the current PPD practices and support the screening process, given the lack of human resources to interview patients. SARs could also be used in other roles, for example to raise social awareness for women to realise that PPD is not uncommon. All whilst ensuring the GP is still in the loop to ensure the safeguard of the patient, as found also in [32]. Accordingly, to PPD experts, a SAR in such a setting could successfully challenge negative gender norms and allow women time to process and realise how they are feeling.

2) Robot's capabilities and appearance: The results point to the need for a number of key skills that a robot used in PPD screening should be endowed with. Given the context of use, i.e., a face-to-face interaction where a robot interviews a patient, it is not surprising that both PPD and gender studies experts highlighted the importance of a robot that can understand what the users are saying and what behaviours they are displaying, and can adapt in a contingent manner. Going further, participants identified a specific set of personalised adaptation skills, which include the ability for the robot to display empathy and mirror low-level user's behaviours, which is key in the design of robots in socially assistive roles [57].

When it comes to the robot's appearance, participants reflected along several dimensions. Not surprisingly, given the context of use, they suggested that the Furhat robot's appearance should be to a great extent human-like, rather than character-like (e.g. manga face). Concerning appearance in gender, participants overall felt that the robot should have feminine traits. While this may carry the risk to perpetuate gender stereotypes, depending on the role played by the robot, patients might feel like talking more to a female looking robot due to an issue of self similarity. In fact, insights from the social sciences suggest that perception of self-other similarity plays a key role in people's ability to relate to others [58]. However, participants were also open to a gender neutral robot, in both appearance and voice. This seems to suggest that the robotic embodiment may offer new opportunities when it comes to robot gendering that can take into account the importance of both perceived self-other similarity, and the need not to perpetuate gender stereotypes.

At a more abstract level, participants reflected on the fact that a robot used in PPD screening will inevitably have to be designed with some values in mind. The latter will affect what role the robot plays, what the robot says, or how it looks, and these design choices might be a reflection of the cultural context, the robot's developers and more in general the norms at the medical institution and the society level. This points to the open question of what should be the values that a robot is designed upon, who should make these decisions and how such decisions can be made in a way that promotes inclusion, non-discrimination and fairness.

This work highlights the importance of a user, expert, and stakeholder-centred approach to the design of robots in PPD screening, which other recent works advocate for [59]. Building on this, we propose that a gender-sensitive intersectional approach is a necessity when it comes to the design of robots for PPD screening, as we show that gender norms surround this sensitive application area for SARs from many

different perspectives. Going even further, results suggest that one may need an approach that not only is gender-sensitive, but intersectionality-sensitive, as often gender aspects cannot be separated by other factors, such as ethnicity.

B. The Future of Governance

1) Three levels of governance: Governance is a wide area, and there are multiple issues to explore and discuss at, at least, three levels. Firstly, there is the soft governance of the (gendered) PPD screening itself, proposed by the national authority. It seems to be struggling with implementation issues in the sense that not all patients (that according to the policy are intended to be screened) are in fact even offered the screening. In addition, the professional informants in our study tend to question the usefulness of the screening tool. Secondly, the established regulatory governance on privacy or integrity poses issues related to the robot as having access to journal data, or data from the former meetings. This could also regard data collecting aspects where the robot is intended to provide the GP or PPD expert with information, which opens for both security and privacy issues [60]. Some participants defined this setting as ethically concerning. Thirdly, which we outline in the section on related work above, the field of AI governance has seen much development lately, especially with regards to fairness issues related to gender, and other, biases.

2) Safeguarding a User-Centric Approach in Healthcare: Indeed, HRI does not have to include an AI-systemic interaction in terms of adaptability and learning, facial recognition or predictive abilities. However, the future social robotics definitely will integrate more AI-methodologies [61]. Therefore, the notions of governance of robots in diagnostic care would benefit from an inclusion of contemporary ideas on AI governance.

Here, there is a focus on the notion of *fairness and anti-discrimination*, including on gender, as a main driver for much policy work, in both the EU [39] as well as globally [44]. Another key theme relates to *transparency*, which in the robot context could relate to how to ensure how patients would know what purpose and role the robot might have, what data it collects (see comments on ethical concern above) and what background training or knowledge it has or has access to. In addition, a closely linked key challenge is *accountability*, that is, there may be a risk of a blurring of who should be held accountable for what, in relation to robotic actions or agency [46]. This leads to an emphasis in European AI governance on *human-centricity* – which could mean the patient – including to keep humans on, or in, the loop – that is, ensuring oversight over the robot.

C. Addressing HRI Issues from an Interdisciplinary Lens

1) Finding Synergies to drive a Gender-Sensitive Approach: The authors met to discuss how gender is accounted for in their respective disciplines. The synergies lied in how SoL could help inform HRI through the study of norms; and how HRI could help SoL make abstract terms come to life through a Furhat robot.

2) A Constant Work-in-Progress to Bridge Disciplines: In every step of the project, at least one author from the field of SoL and HRI were present. For example, after each interview the first and second author would discuss how they perceived the interview and what remarks they found important. This exercise allowed for a constant bridging of disciplines. This exercise was later important when transitioning into coding, as the authors were both aware of what the other discipline would seek in the analysis. Furthermore, we discovered that it is crucial to continuously inform each other on the practices and perceptions of each other's fields to find common ground.

VI. Conclusion

In this study we have investigated how SARs, designed according to the principle of gender fairness, can impact future practices of governance in the medical institution. Through a vignette-informed interview study with PPD professionals and gender scholars, this works contributes to notions and knowledge of gender fairness in the HRI domain. We have pointed to what role gender norms play for a SAR-assisted PPD-screening and how current medical practices can be challenged or, likely, developed.

The findings point towards the PPD experts being willing to bring robots into the PPD setting. The PPD experts saw the robot as an integration to their work within their medical team. On the other hand, gender experts were much more critical of the SARs in this medical setting, pointing to macro-level issues emanating from the medical institution itself. When it comes to HRI design implications, we derive a number of recommendations on SARs' possible contexts of use within and outside of the medical institution, roles they could play, as well as physical characteristics and capabilities.

Lastly, we see three layers of governance aspects, from the PPD screening explicitly, to privacy issues with data handling robots in diagnostic care, to the ethically framed guidelines of contemporary AI governance. These emphasise the importance of human-centricity, oversight and user autonomy and transparency, which clearly relate to already established medical ethics emphasising patient autonomy and right to information.

VII. Limitations and Future Work

This work was limited to experts in the Swedish context; thus the results cannot be generalised to a wider context but can help incorporate gender-sensitive approaches to HRI design. Furthermore, studies on users specifically are needed to understand their perspective on SARs and if SARs would fit the medical sector. Finally, we would suggest widening the stakeholders to more than just the pregnant women, and include their partner(s) at the very least.

VIII. Acknowledgement

This work was supported by (i) the Wallenberg AI, Autonomous Systems and Software Program—Humanities and Society (WASP-HS) funded by the Marianne and Marcus Wallenberg Foundation and the Marcus and Amalia Wallenberg Foundation, and (ii) the WoMHeR Centre, Uppsala University, Sweden.

REFERENCES

[1] C. Woody, A. Ferrari, D. Siskind, H. Whiteford, and M. Harris, "A systematic review and meta-regression of the prevalence and incidence of perinatal depression," *Journal of affective disorders*, vol. 219, pp. 86–92, 2017.

[2] D. Roussos-Ross, "Perinatal depression," pp. 24–36, aug 2019.

[3] S. M. Marcus and J. E. Heringhausen, "Depression in childbearing women: when depression complicates pregnancy," *Primary Care: Clinics in Office Practice*, vol. 36, no. 1, pp. 151–165, 2009.

[4] E. Q. Cox, N. A. Sowa, S. E. Meltzer-Brody, and B. N. Gaynes, "The perinatal depression treatment cascade," *Journal of Clinical Psychiatry*, vol. 77, no. 09, p. 1189–1200, 2016.

[5] L. D. Riek, *Robotics Technology in Mental Health Care*. Elsevier, 2016, p. 185–203.

[6] B. Scassellati and M. Vázquez, "The potential of socially assistive robots during infectious disease outbreaks," *Science Robotics*, vol. 5, no. 44, p. eabc9014, 2020.

[7] S. Jeong, S. Alghowinem, L. Aymerich-Franch, K. Arias, R. Lapedriza, A. andPicard, H. Park, and C. Breazeal, "A robotic positive psychology coach to improve college students' wellbeing," in *RO-MAN 2020*. IEEE, 2020.

[8] C. Criado Perez, *Invisible Women: Data Bias in a World Designed for Men*. Penguin Random House, 2020.

[9] S. Law, I. Ormel, S. Babinski, D. Plett, E. Dionne, H. Schwartz, and L. Rozmovits, "Dread and solace: Talking about perinatal mental health," *International Journal of Mental Health Nursing*, p. inm.12884, May 2021.

[10] A. Ostrowski, C. Breazeal, and H. Park, "Design research in hri: Roboticists, design features, and users as co-designers," in *RO-MAN 2020, Workshop on Designerly HRI Knowledge*. IEEE, 2020.

[11] AIHLEG, "Ethics guidelines for trustworthy AI," 2019.

[12] WHO, *Ethics and Governance of Artificial Intelligence for Health: WHO guidance*, who guidan ed. World Health Organisation, 2021. [Online]. Available: http://apps.who.int/bookorders.

[13] World Health Organisation, "Fact sheet: Depression," 2020. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/depression

[14] A. J. Ballantyne and W. A. Rogers, "Sex bias in studies selected for clinical guidelines," *Journal of Women's Health*, vol. 20, no. 9, pp. 1297–1306, 2011.

[15] J. A. Brommelhoff, K. Conway, K. Merikangas, and B. R. Levy, "Higher rates of depression in women: Role of gender bias within the family," *Journal of Women's Health*, vol. 13, no. 1, pp. 69–76, 2004.

[16] A. Bacigalupe and U. Martín, "Gender inequalities in depression/anxiety and the consumption of psychotropic drugs: Are we medicalising women's mental health?" *Scandinavian Journal of Public Health*, vol. 49, no. 3, pp. 317–324, 2021.

[17] N. A. Atwood, "Gender bias in families and its clinical implications for women." pp. 23–36, 2001.

[18] B. Wickberg, M. Bendix, M. B. Wetterholm, and A. Skalkidou, "Perinatal mental health around the world: priorities for research and service development in sweden," *BJPsych International*, vol. 17, no. 1, pp. 6–8, 2020.

[19] Socialstyrelsen, "Vård efter förlossning. En nationell kartläggning av vården till kvinnor efter förlossning," 2017. [Online]. Available: https://www.socialstyrelsen.se/globalassets/sharepoint-dokument/artikelkatalog/ovrigt/2017-4-13.pdf

[20] S. Varrasi, S. Di Nuovo, D. Conti, and A. Di Nuovo, *A Social Robot for Cognitive Assessment*, 2018.

[21] L. D. Riek, "Healthcare robotics," *Communications of the ACM*, vol. 60, no. 11, pp. 68–78, 2017.

[22] A. N. Joinson, "Self-disclosure in computer-mediated communication: The role of self-awareness and visual anonymity," *Eur. J. of Social Psych.*, vol. 31, no. 2, p. 177–192, 2001.

[23] R. E. Simut, J. Vanderfaeillie, A. Peca, G. Van de Perre, and B. Vanderborght, "Children with autism spectrum disorders make a fruit salad with probo, the social robot: an interaction study," *Journal of autism and developmental disorders*, vol. 46, no. 1, pp. 113–126, 2016.

[24] M.-T. Chu, R. Khosla, S. M. S. Khaksar, and K. Nguyen, "Service innovation through social robot engagement to improve dementia care quality," *Assistive Technology*, vol. 29, no. 1, pp. 8–18, 2017.

[25] S. Varrasi, S. Di Nuovo, D. Conti, and A. Di Nuovo, "A social robot for cognitive assessment," in *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 269–270. [Online]. Available: https://doi.org/10.1145/3173386.3176995

[26] M. A. Saleh, F. A. Hanapiah, and H. Hashim, "Robot applications for autism: a comprehensive review," *Disability and Rehabilitation: Assistive Technology*, vol. 16, no. 6, pp. 580–602, 2021.

[27] E. E. Lee, J. Torous, M. D. Choudhury, C. A. Depp, S. A. Graham, H.-C. Kim, M. P. Paulus, J. H. Krystal, and D. V. Jeste, "Artificial intelligence for mental health care: Clinical applications, barriers, facilitators, and artificial wisdom," *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, vol. 6, no. 9, p. 856–864, 2021.

[28] D. Conti, S. Di Nuovo, S. Buono, and A. Di Nuovo, "Robots in education and care of children with developmental disabilities: a study on acceptance by experienced and future professionals," *International Journal of Social Robotics*, vol. 9, no. 1, pp. 51–62, 2017.

[29] D. DeVault, R. Artstein, G. Benn, T. Dey, E. Fast, A. Gainer, K. Georgila, J. Gratch, A. Hartholt, M. Lhommet, G. Lucas, S. Marsella, F. Morbini, A. Nazarian, S. Scherer, G. Stratou, A. Suri, D. Traum, R. Wood, and L.-P. Morency, "Simsensei kiosk: A virtual human interviewer for healthcare decision support," vol. 2, 01 2014, pp. 1061–1068.

[30] D. Suendermann-Oeft, A. Robinson, A. Cornish, D. Habberstad, D. Pautler, D. Schnelle-Walka, F. Haller, J. Liscombe, M. Neumann, M. Merrill *et al.*, "Nemsi: A multimodal dialog system for screening of neurological or mental conditions," in *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*, 2019, pp. 245–247.

[31] "A roadmap for US robotics: From internet to robotics 2020," *Computing Community Consortium*, 2020.

[32] M. Zhong, A. M. Bilal, F. C. Papadopoulos, and G. Castellano, "Psychiatrists' views on robot-assisted diagnostics of peripartum depression," in *Social Robotics*, H. Li, S. S. Ge, Y. Wu, A. Wykowska, H. He, X. Liu, D. Li, and J. Perez-Osorio, Eds. Springer International Publishing, 2021, p. 464–474.

[33] A. Weiss and K. Spiel, "Should robots have a gender — should they be gendered?" in *Proceedings of the GenR Workshop - Gendering Robots: Ongoing (Re)configurations of Gender in Robotics, IEEE RO-MAN*, 2021.

[34] M. Weßel, N. Ellerich-Groppe, and M. Schweda, "Gender stereotyping of robotic systems in eldercare: An exploratory analysis of ethical problems and possible solutions," *International Journal of Social Robotics*, pp. 1–14, 2021.

[35] M. West, R. Kraut, and H. Ei Chew, "I'd blush if i could: closing gender divides in digital skills through education," 2019.

[36] D. Bryant, J. Borenstein, and A. Howard, "Why should we gender? the effect of robot gendering and occupational stereotypes on human trust and perceived competency," in *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 13–21. [Online]. Available: https://doi.org/10.1145/3319502.3374778

[37] R. B. Jackson, T. Williams, and N. Smith, "Exploring the role of gender in perceptions of robotic noncompliance," in *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 559–567. [Online]. Available: https://doi.org/10.1145/3319502.3374831

[38] K. Winkle, G. I. Melsión, D. McMillan, and I. Leite, "Boosting robot credibility and challenging gender norms in responding to abusive behaviour: A case for feminist robots," in *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '21 Companion. New York, NY, USA: Association for Computing Machinery, 2021, p. 29–37. [Online]. Available: https://doi.org/10.1145/3434074.3446910

[39] S. Larsson, "On the governance of artificial intelligence through ethics guidelines," *Asian Journal of Law and Society*, vol. 7, no. 3, pp. 437–451, 2020.

[40] A. Taeihagh, "Governance of artificial intelligence," *Policy and Society*, pp. 1–21, 2021.

[41] S. Y. Tan and A. Taeihagh, "Governing the adoption of robotics and autonomous systems in long-term care in singapore," *Policy and society*, pp. 1–21, 2020.

[42] C. Högberg and S. Larsson, "Ai and patients' rights: Transparency and information flows as situated principles in public health care," in *De Lege årsbok 2021: Law, AI & Digitalization*. Uppsala, Sweden: Iustus, 2022.

[43] A. Shaban-Nejad, M. Michalowski, and D. L. Buckeridge, *Explainable AI in Healthcare and Medicine: Building a Culture of Transparency and Accountability*. Springer Nature, 2021, vol. 914.

[44] A. Jobin, M. Ienca, and E. Vayena, "The global landscape of ai ethics guidelines," *Nature Machine Intelligence*, vol. 1, no. 9, pp. 389–399, 2019.
[45] T. Hagendorff, "The ethics of ai ethics: An evaluation of guidelines," *Minds and Machines*, vol. 30, no. 1, pp. 99–120, 2020.
[46] W. H. Organization *et al.*, "Ethics and governance of artificial intelligence for health: Who guidance," 2021.
[47] S. Larsson, "Ai in the eu: Ethical guidelines as a governance tool," *The European Union and the Technology Shift*, pp. 85–111, 2021.
[48] M. Ebers, V. R. Hoch, F. Rosenkranz, H. Ruschemeier, and B. Steinrötter, "The european commission's proposal for an artificial intelligence act—a critical assessment by members of the robotics and ai law society (rails)," *J*, vol. 4, no. 4, pp. 589–603, 2021.
[49] J. Mökander, M. Axente, F. Casolari, and L. Floridi, "Conformity assessments and post-market monitoring: a guide to the role of auditing in the proposed european ai regulation," *Minds and Machines*, pp. 1–28, 2021.
[50] M. Veale and F. Z. Borgesius, "Demystifying the draft eu artificial intelligence act—analysing the good, the bad, and the unclear elements of the proposed approach," *Computer Law Review International*, vol. 22, no. 4, pp. 97–112, 2021.
[51] R. Benjamin, *Race After Technology: Abolitionist Tools for the New Jim Code*. Polity, 2019.
[52] C. D'Ignazio and L. F. Klein, *Data Feminism*. The MIT Press, 2020.
[53] N. Arista, S. Costanza-Chock, V. Ghazavi, and S. Kite, *Against Reduction: Designing a Human Future with Machines*. MIT Press, 2021.
[54] K. W. Crenshaw, "Race, reform, and retrenchment: Transformation and legitimation in antidiscrimination law," *Harv. L. rev.*, vol. 101, p. 1331, 1987.
[55] S. Serholt *et al.*, "The case of classroom robots: teachers' deliberations on the ethical tensions," *AI and Society*, vol. 32, no. 4, pp. 613–631, 2017.
[56] G. Guest, K. M. MacQueen, and E. E. Namey, *Applied Thematic Analysis*. SAGE, Dec 2014. [Online]. Available: https://methods.sagepub.com/book/applied-thematic-analysis
[57] P. Alves-Oliveira, P. Sequeira, F. Melo, G. Castellano, and A. Paiva, "Empathic robot for group learning: A field study," *ACM Transactions in Human-Robot Interaction*, vol. 8, no. 1, 2019.
[58] P. Bourgeois and U. Hess, "The impact of social context on mimicry," *Biological Psychology*, vol. 77, pp. 343–352, 2008.
[59] A. Ostrowski, C. Breazeal, and H. Park, "Long-term co-design guidelines: Empowering older adults as co-designers of social robots," in *RO-MAN 2021*. IEEE, 2021.
[60] C. Lutz, M. Schöttler, and C. P. Hoffmann, "The privacy implications of social robots: Scoping review and expert interviews," *Mobile Media & Communication*, vol. 7, no. 3, pp. 412–434, 2019.
[61] M. E. Foster, S. Ali, S. Litwin, J. Parker, R. P. Petrick, D. H. Smith, J. Stinson, and F. Zeller, "Using ai-enhanced social robots to improve children's healthcare experiences," in *International Conference on Social Robotics*. Springer, 2020, pp. 542–553.