

University of Tartu
Institute of Philosophy and Semiotics

THE ETHICAL DESIGN OF SEX ROBOTS

Master's Thesis in Philosophy

Ezgi Mizrak

Supervisor: Francesco Orsi

Number of characters: 90993

Tartu, 2024

Acknowledgements

I would first like to express my gratitude to my supervisor, Francesco Orsi. Over the last two years, he has been an excellent mentor, patiently guiding me from the basics of writing to finishing this thesis.

I am also deeply thankful to my mother, Belma Şanlı, who has always allowed me to be myself and respected my choices, even when the topic of sex robots initially gave her pause. I am deeply thankful to my dear friends, my chosen family in Turkey, my friends in Estonia, and Kosmos the dog. They have made me more resilient and encouraged me to complete this thesis.

TABLE OF CONTENTS

I. INTRODUCTION	4
1.1. Defining Sex Robots	4
1.2. Technological Development of Sex Robots	5
1.3 Overview of the Thesis	6
II. THE ETHICS OF SEX ROBOTS: FROM THE SYMBOLIC CONSEQUENCES ARGUMENT TO THE HARM REDUCTION MODEL	8
2.1. Which Ethical Theory to Evaluate Sex Robots?	8
2.2. A Consequentialist Argument Against Sex Robots: The Symbolic Consequences	10
2.3. Is the Symbolic Consequences Argument the Last Word on the Ethics of Sex Robots?	14
2.3.1 Intrinsically Harmless Enjoyment	14
2.3.2 Love/Companionship for The User	15
2.3.3 Therapeutic Benefits	16
2.4. How to Move Forward? Three Approaches	17
2.4.1 Full Prohibition Model	18
2.4.2 Full Unregulated Model	21
2.4.3 In Favor of the Harm Reduction Model	21
2.5 Summary of Chapter II	23
III. ETHICAL DESIGN: CONSENT	24
3.1. Introduction	24
3.2 What is Consent?	24
3.3 Should Sex Robots Have a Consent Module? Sparrow's Arguments	25
3.4 The Appropriate Consent Module	28
3.5 How to Design a Consent Module Ethically?	29
3.5.1 Minimizing Risks: Ethical Design of Consent Simulation	30
3.5.2 Making Consent Clear: Scripting Ethical Interactions	34
3.6 Summary of Chapter III	37
IV. CONCLUSION	39
Abstract	41
REFERENCES	42
APPENDIX	46

I. INTRODUCTION

Sex robots bring significant ethical concerns as their use and development increase. These ethical concerns primarily focus on what sex robots symbolize and the potential consequences of this symbolism in human relationships and society in general. In this thesis, I will argue, first, that while there are serious concerns about sex robots, we should adopt a harm reduction model rather than a full prohibition or a fully unregulated model. Second, I will argue that the potential harm of sex robots can be reduced and mitigated by ethically regulating their design, in particular by way of including an appropriate consent module.

In this first chapter, first, I will define what a sex robot is and how it differs from other sexual devices. Then, I will explain the progress of the sex robot models developed and understand what technology might offer in the future.

1.1. Defining Sex Robots

Sex robots are inanimate sexual tools designed with advanced artificial intelligence and robotic materials to provide a sexual experience similar to that of humans. While basic sex toys like dildos, vibrators, and love dolls are static and offer physical stimulation without interaction or response, sex robots are designed to mimic human behavior and responsiveness. They are meant to simulate the humanoid sexual experience more realistically. These robots have features ranging from voice recognition to moving limbs and facial expressions. So far, they can have simple conversations and respond to physical touch. Although the features of current sex robots are limited, they are predicted to eventually approach the level of real human interaction (Levy, 2007, p. 303). This is important for my thesis because I will point out that important ethical issues raised by sex robots are related to their symbolism, and the degree of how closely they resemble humans. Second, this thesis makes specific proposals on how to design sex robots in such a way as to mitigate the negative symbolic consequences. Therefore, how similar these robots are to humans, how indistinguishable they may become in the future, and who and what they represent are crucial. Even though it is clear that they are robots, their similarity to humans is an important consideration because of their symbolic consequences for society. Considering sex robots that closely resemble humans, the influence on symbolic consequences of these robots on society would be strong. Conversely, if robots behave in a strongly robotic manner, symbolic consequences of these robots on society are less likely. Consequently, an addition to the definition of the current understanding

of sex robots is warranted: Sex robots are tools that can reflect the appearance and behavior of humans (particularly that of women) to varying degrees according to certain patterns, meaning they could potentially symbolize humans in an accurate or distorted way.

1.2. Technological Development of Sex Robots

As stated in the definition of sex robots, today's sex robots are limited in what they can do and how they look in comparing to humanlike features. However, examining the progress of the robot models developed so far, - compared to traditional sex dolls-, can give us an idea of what future technology might offer and its impact on society. Thus, we seek to demonstrate that the concerns are not born out of an undue exaggeration or a sense of despair about a dystopian future, but are significant issues deserving of detailed ethical examination, given the current societal issues concerning gender, sexuality, and relationships.

To have a more concrete idea of what sex robots are now (2024) and what they may be in the future, let us consider the world's first sex robot Roxxy, manufactured in 2010, and compare it to today's late advancement of sex robot Harmony: a sophisticated humanoid robot designed for interactive human engagement. Harmony is equipped with advanced AI, enabling learning and adapting to human behaviors. This robot features customizable physical appearances, including eye color, hairstyle, makeup style, and skin tone, and stands approximately 170cm. The RealDollx body of Harmony is made from metal and silicone. The manufacturer claims that the skin is soft and tactile, as well as the durability and structural integrity of the metal frame, which enhances Harmony's realistic appearance. (RealDoll X, 2023). The ad on the company's website includes a detailed description suggesting that the robot can express arousal on its face when touched through its genital area. This is made possible by "The Sensex, an electronic insert designed to pair via Bluetooth with our X-Mode App. It can detect touch, movement, and transitions from mild arousal to orgasm" (RealDoll X, 2023). Users interact with the robot by speaking or typing commands through the RealDoll X smartphone app. When the user is talking, the app uses speech recognition to interpret and process the input. As a result, the robot then responds to these inputs using its built-in speakers.

As of now, we can see from the example of Harmony that sex robots have their way to lifelikeness already. This is especially clear when compared to the first sex robot ever, Roxxy,

introduced by TrueCompanion in 2010. It features more basic AI capabilities, with limited conversational interactions based on pre-programmed responses. It had fewer choices regarding appearances and also looked more like a classic sex doll (Contributor, 2010).

As RealdollX company claims, it is possible that sex robots will be even more advanced than products like Harmony : "They will soon have hyper-realistic features, warmth, and sensors that react to touch. When interacting with these dolls, we want users to ask themselves, 'What is she thinking?'"(RealDoll X, 2023).

Finally, according to (Bedbible Research Center, 2024), approximately 17.4% of people have had sex with a sex robot. Perhaps surprisingly, the gender differences are not much in owning sex robots, men (17.8%) and women (16.5%). Globally, the sex robot market is worth \$200.7 million annually, a small fraction of the \$37 billion sex toy industry. On average, 156 sex robots are sold per day, and the prices of sex robots decreased, starting from \$24,000 in 2010 to \$3,567 in 2022. These data give reason to think that access to sex robots will only increase in the foreseeable future.

1.3 Overview of the Thesis

In Chapter II, I will first discuss which ethical theory is the sounder option for evaluating sex robots. I will argue that utilitarianism can provide a better starting point than deontology or virtue ethics because utilitarian reasons for or against sex robots have to be assessed by these other theories as well. Then, I will present in detail one particular consequentialist argument against sex robots: the Symbolic Consequences Argument. On the other hand, I will point out the benefits of sex robots. Following this, I will argue that a harm reduction model is the most ethically sound option for future regulation of sex robots because the benefits of sex robots are greater than the harm they could cause with ethically sound design regulations.

In Chapter III, I will first tackle one aspect of the ethical design of sex robots: whether sex robots should or should not include a consent module. Next, I will define consent and how it is considered in human relationships, then discuss the ethical importance of sexual consent and why it is ethically valuable and essential. I then explain the conditions under which consent is valid and

discuss the importance of sexual consent due to the value of bodily autonomy. Next, I will explore what happens when consent is not present in sex robots and argue that implementing a consent module in sex robots is a more ethically sound design choice from a utilitarian perspective, particularly to minimize harm to women, due to the possible contribution of robots to promoting rape culture. To do this, I will first show how consent simulation can be ethically designed and provide examples of simulated consent scenarios for ethical design choices.

Finally, in Chapter IV, I conclude by presenting the aim of the thesis, summarizing its content and limitations, and emphasizing the need for further empirical research on the ethics of sex robots.

II. THE ETHICS OF SEX ROBOTS: FROM THE SYMBOLIC CONSEQUENCES ARGUMENT TO THE HARM REDUCTION MODEL

2.1. Which Ethical Theory to Evaluate Sex Robots?

In this thesis, I will adopt a broadly utilitarian approach to evaluate the ethics of sex robots. This is the approach used by philosophers like John Danaher but is also broadly found in critics of sex robots, such as Kathleen Richardson. To do this, I will first explain utilitarianism and the consequentialist arguments. Subsequently, I will explain non-consequentialist theories such as deontology and virtue ethics. Ultimately, I will explain why utilitarianism is a more comprehensive theory compared to other theories for the ethics of sex robots.

Utilitarianism, as a consequentialist theory approach that determines morality according to the consequences of actions, differs from other moral theories such as deontology and virtue ethics. According to this view, the aim is to maximize general happiness and minimize suffering and harm. John Stuart Mill summarizes this as "actions are right in proportion as they tend to promote happiness; wrong as they tend to produce the reverse of happiness"(Mill, 1863).

Act utilitarianism, one of two variations of utilitarianism, evaluates each action by its direct consequences, while rule utilitarianism generally focuses on rules that maximize utility when followed. To explain these two approaches with an example, for act utilitarians, lying can be justified if it leads to more happiness and in contrast, rule utilitarians would say that lying is wrong, even when it may lead to greater happiness, because following the rule "do not lie (even when this maximizes happiness)" is more beneficial in general (Nathanson, 2014).

In this thesis I will remain neutral between act utilitarianism and rule utilitarianism. However, since my topic is how it is best for society to regulate production and use of sex robots, my primary focus will be on which rules are best, from a utilitarian point of view, for society to follow in this matter - rather than on whether a given single user's act of buying or using a sex robot is morally right or wrong.

I will also remain neutral on whether "happiness" should be understood in hedonistic terms, as pleasure, or instead in terms of desire, as the satisfaction of preferences. When it comes to evaluating sex robots, pleasure (and pain) is clearly important, but (as we will see below) the good or bad consequences of sex robots can have an impact on people that goes beyond mere sensations

and affects their overall life satisfaction or frustration. Therefore, I will not choose between hedonistic utilitarianism and preference utilitarianism.

Deontology, conversely, argues that actions are morally right or wrong based on a set of rules, regardless of the consequence (Alexander & Moore, 2021). These principles, like honesty and justice, may impose absolute or only *prima facie* duties, but in any case, imply that it is not always right to maximize happiness or other goods. In the example above, the act of lying would be wrong because it violates the principle 'do not lie', regardless of whether following this principle is something that maximizes total happiness. On the other hand, virtue ethics, which has its roots in Ancient Greece, focuses on developing virtuous characteristics of the individual agent and making decisions that reflect these virtues rather than on specific actions or their consequences (Athanasoulis, n.d.).

It can be seen that moral decision mechanisms could be shaped on several grounds: act utilitarianism uses harm-benefit analysis case by case, rule utilitarianism appeals to beneficial rules, deontology follows moral duties and virtue ethics considers actions that improve moral character.

However, it is essential to recognize that the ethical concerns of utilitarians must also be shared by any plausible version of deontology and virtue ethics, particularly regarding the consequences for overall happiness. For instance, a deontological thinker might include, alongside honesty and justice, a principle such as "maximize benefits and minimize harms" and, in the end, reflect a concern for the outcomes of actions. Similarly, virtue ethics can consist of the virtues of beneficence and non-maleficence which relate to promoting happiness and minimizing harm. Nevertheless, these approaches have limitations compared to utilitarianism when it comes to evaluating the ethics of sex robots. By focusing strictly on rules, deontology might ignore the practical consequences of using sex robots. This could lead to morally rigid positions that do not adequately address the potential benefits, such as reducing loneliness or providing safe sexual experiences for marginalized individuals. Moreover, virtue ethics, as discussed by philosophers like (Peeters & Haselager, 2021) in the context of sex robots, focuses on the virtues or moral character developed through actions. Peeters and Haselager (2021) suggest that applying virtue ethics for sex robots could support developing or strengthening compassion in supervised therapeutic scenarios (Peeters & Haselager, 2021, p. 63). This proposition may be accurate, but it can also be argued virtue ethics being so centered on the agent, may not be a general adequate

ground for deciding what to do about sex robots regarding the negative consequences at large that may arise in society.

Finally, utilitarianism can adapt to new and evolving situations by assessing the specific outcomes of using sex robots in different contexts. This flexibility allows for a more comprehensive understanding of sex robots. Since utilitarianism requires considering the interests of all affected parties equally, sex robots can be considered for both impacts on users and society, thus leading to more inclusive and equitable ethical decisions.

2.2. A Consequentialist Argument Against Sex Robots: The Symbolic Consequences

In this section, I will explain one of the main arguments in the debate over the ethics of sex robots, which is the symbolic consequences argument. To start with a definition, the symbolic consequences argument presents a consequentialist critique based on the symbolic impact of sex robots on societal norms and behaviors. This argument focuses on a specific type of negative consequence: the perpetuation of harmful stereotypes and attitudes through the symbolic representation of gender and sexuality in robots.

I will start with the most cited structure of the argument, which is presented by John Danaher, and then expand on his argument. Danaher puts the argument in the following way:

(1) Sex robots do/will symbolically represent ethically problematic sexual norms. (Symbolic Claim.) (2) If sex robots do/will symbolically represent ethically problematic sexual norms, then their development and/or use will have negative consequences. (Consequential Claim.) (3) Therefore, the development and/or use of sex robots will have negative consequences and we should probably do something about this. (Warning Call Conclusion). (Danaher, 2017, p. 107)

I will now reconstruct the argument step by step, explaining each claim according to the premises below. In my reconstruction, the two premises in Danaher's argument (Symbolic Claim and Consequential Claim) will be shown to be themselves the conclusions of two mini-arguments.

The argument goes as follows:

1. Sex robots represent humans (particularly women) through appearance and behavior.
2. The representation of humans in sex robots follows the blueprint of representation of women in mainstream pornography. (Disregard for consent-seeking steps, having submissive behaviors, maintaining traditional body types, and "coquettish manner"(Danaher, 2017, p. 107).
3. The way women are represented in porn is not ethical.

4. If 3, then the way women are represented in sex robots is not ethical (See Symbolic Claim).
5. With the use of sex robots, the line between women and the representation of women will blur for the users.
6. If 5, then the user will equate women with their representations in porn or sex robots and exhibit harmful behavior accordingly (see Consequential Claim).

(C) “Therefore, the development and/or use of sex robots will have negative consequences and we should probably do something about this. (Warning Call Conclusion)” (Danaher, 2017, p. 107)

As mentioned before in Chapter I, the first and second premise of this argument can be clearly seen in today's market already. The company RealDollX offers a category called Wicked, featuring a collection of sex robots that resemble actual porn actresses. Their advertisement states: “Bring Your Favorite Porn Star Home Our Wicked RealDolls are the perfect way to make your fantasy a reality. Each doll is lifelike and made true to the porn model’s actual measurements. Spend a night with your dream girl whenever you feel like it!” (RealDoll X, n.d.)

Regarding second premise, most of the robots that are in market right now have a certain style of forced appearance that can be called as “porn-star-esque features” (Danaher, 2017, p. 107). This appearance is not necessarily something evil that should not be seen in a sex robot's design, as with other body types. Instead, it is criticized by the proponents of the symbolic consequences argument because it becomes an end in the mainstream understanding of what is 'ideal' in the male gaze.

What is more interesting is once the user has the body that they like. “All faces are compatible with all body types. However, some faces and bodies will look better together than others due to differences in size.”(RealDoll X, 2023) Furthermore, the user can choose what kind of labia they want or if they want pubic hair or transgender options. This basically means that the robot can be adapted to whatever the user desires and be utilized in any way the user prefers. This implies that within the example of the sex robots of RealDoll Company, market at hand develops robots that are symbolically problematic in the context of appearance and behaviors. Despite the apparently large variety of choices for users, the reality is that these options are all inspired by pornographic representations of women.

It is challenging to prove the Consequential Claim (premises 5,6) due to the lack of data on user behavior, even if the Symbolic Claim may be granted. The Consequential Claim is about the transfer of behaviors and values acquired after interaction with sex robots from robot partner to human partner, that is, to real-life relationships. However, Richardson, one of the greatest proponents of the Symbolic Consequences argument, has not detailed this symbolic transfer process, nor has anyone else. (Bisconti, 2021) uses an analogy to prove that the feared symbolic transfer need not necessarily happen: theatre. Just as an actor portraying a murderer does not become a murderer or show increased aggression, those who watch non-consensual violence or immorality in games do not internalize these behaviors. He claims actors and audiences can distinguish between fictional representations and real-life action:

If an actor on stage impersonates an evil murderer, even though repeating this part every day for weeks, hours and hours a day, even though a certain degree of identification is required for a convincing actor performance, he/she will not become a murderer, nor will visibly increase its aggression in real life. Likewise, viewers who look at the representation of a non-consensual violence, rape, immoralities of various kinds, will not introject this symbolism into their social rules. So, the person who performs the actions, the actor, will not transfer the behavioural patterns out of the fictional context of the play; viewers will not assimilate a symbolic universe made of abuse and immorality. Why does the shift not happen in this case? (Bisconti, 2021, p. 563).

It may be expected that also sexual interactions with sex robots will remain within their intended context. However, this does not mean that the concern about the Consequential Claim is in vain. Indeed, we do not carry our behaviors outside their context every time, and there are many examples of this. But when it comes to sexism and violent behaviors, the pattern changes. Bisconti's argument, which seems very sound at first glance, incorrectly weighs the representation of acts and the real acts. An actor can play a murderer daily and show no signs of being affected. In the state of acting, the situation goes as follows: (i) the actor does play, (ii) there is a set scenario (iii) the actor has an audience. First, the actor is aware that he is performing an act, he has an audience, and he plays to his audience on a stage; if it is not an improvised act, the whole scenario is clear, and there is a certain beginning and end. Now let's think about the same thing again with a person using a sex robot: (i) the user does not play, (ii) there is no set scenario, the user acts according to his desires and emotions, (iii) the user does not have an audience if not preferred. While the actor plays the killer, he pretends to kill the victim, but the sex robot user does not

pretend to have sex with the sex robot; he engages in sexual interaction and does it as if he were doing it with a real person. The player represents the murderer and is unaffected by what he represents, which is true, but there is a different relationship between the sex robot and the user: the sex robot represents a human being, and the user uses this representation. Here, we can give a perception from the audience: the audience is exposed to what is represented but still can distinguish between reality and play, and this is also true. Because what happens there is that a play is a play, but a sexual interaction with the sex robot is real sexual interaction. The issue is not that the sex robot is not a real person; the issue is the reality of the action performed. In the first, the action is not real but a play, while in the latter, the action is real.

Now, we can consider same analogy with pornography and a specific kind of audience, teenagers. One study shows that male adolescents exposed to violent pornography are 2-3 times more likely to report sexual TDV (Teen Dating Violence) perpetration, victimization, and physical TDV victimization compared to their non-exposed peers. Female adolescents exposed to the same content are 1.5 times more likely to experience TDV compared to their unexposed peers (Rostad et al., 2019). Even though we can see that violent pornography significantly affects both boys and girls, Rostad and colleagues quotes a report that "although a considerable portion of participants reported exposure to pornography depicting rape or sexual violence at least once before age 18, significantly more boys reported exposure than girls" (Rostad et al., 2019, p. 2139).

Rostad and colleagues observe:

Exposure to violent pornography may relate to TDV to a greater extent for boys because they may be more frequently exposed and exposed to more extreme images, which may contribute to attitudes more accepting of interpersonal violence, and sexual violence in particular. On the other hand, some male adolescents may simply possess a propensity for aggression, which is manifested in violence in their dating relationships and intentional consumption of images that depict violence and degradation. (Rostad et al., 2019, p. 2144)

Considering teenagers exposed to violent pornography, one might argue that these portrayals are merely performances, similar to plays, and not reflective of reality. However, this comparison is flawed. Unlike in theaters or movies where scenarios are purely fictional, the sex acts performed by actors in pornography involve real physical interactions. Therefore, the impact on viewers may differ significantly from that of traditional media. We can see that the line between the learned behaviors from pornography and the behaviors of real-life dating experiences blur, as

stated in premise 6 pornography as a form of representation of women affects the behaviors towards women.

Thus far, I explained that sex robots indeed are a symbol of ethically problematic sexual norms. Second, I explained that although it is difficult to prove that there will be consequences of this symbolism in society, with the reasoning that shows the correlation between pornography and violence, it is not difficult to predict the consequences of using and developing sex robots. Therefore, it is imperative to do something to address the issues tied to the use and development of sex robots. Based on the arguments presented, what Danaher calls the "Warning Call Conclusion" is debated differently by scholars. With the lack of data on use of sex robots and the societal shift it may cause, the exact measures to be taken are still unclear. I come back to this in section 2.4, "Three Approaches".

2.3. Is the Symbolic Consequences Argument the Last Word on the Ethics of Sex Robots?

As can be understood from the Symbolic Consequences Argument, the use and development of sex robots may have potentially negative consequences and the warning call about doing something about these consequences before they rise is not in vain. However, the claims made in this argument are not absolute. Ultimately, the design and production choices of the manufacturer are not dictated by nature and are always subject to change. As Danaher asks, "Are the design, manufacture, sale, and use of sex robots likely to reinforce and exacerbate the problematic symbolism? Or could these factors have positive consequences that are capable of outweighing (and thus warranting changes in) the symbolic interpretation?" (Danaher, 2017, p. 118). It is reasonable to assume that sex robots could offer potential benefits that are not entirely unrealistic. In an ideal world, with the needed changes that are made to the design, manufacture, sale, and use of sex robots, we can simply imagine the benefits these robots would have on humanity at large. I will discuss these potential benefits from three perspectives in the following sections: first, harmless enjoyment for the user; second, love/companionship for the user; and third, therapeutic benefits.

2.3.1 Intrinsically Harmless Enjoyment

When we think about the benefits of sex robots, the most prominent feature is that they make it possible to provide direct sexual pleasure, and as a result, the amount of happiness will

increase. To illustrate the potential utility of sex robots, consider their application among various groups such as women, individuals who have not publicly disclosed their sexual orientation, and those who are otherwise unable to engage in sexual activities. The possibility of more humanoid robots in the future promises us that sexual intercourse with a robot can be similar to the sexual intercourse that can be experienced with a human, and thus promises that there will be sexually satisfied and happier individuals. Neil McArthur references a study where it appears that for the average person, increasing the frequency of sexual intercourse from once a month to at least once a week provides as much additional happiness as a \$50,000 annual salary increase. Higher levels of sexual satisfaction not only increase people's perceived happiness levels but also help them become healthier. High levels of sexual activity have been associated with weight loss, lower stress levels, better heart and blood pressure outcomes, lower rates of prostate cancer in men, and better sleep. Some of these benefits of sexual satisfaction are made possible by sex with a partner or solo sex. But others are the result of the physical exertion required to have sex with a partner and the psychological well-being that results from sex with a partner. According to McArthur, one study has shown that sex with a partner has some psychological benefits that masturbation cannot provide (McArthur, 2017, p. 34). This means that sex with a robot, even if it does not replace sex with another human, can have better psychological benefits than masturbation or simply being alone while engaging in sexual activity.

To the extent that robots do not suffer, and the interaction remains healthy for the user (for example, it does not decrease his happiness in his relationships with actual people), sex robots seem to be a source of intrinsically harmless enjoyment, especially for many people who would otherwise have difficult access to sexual satisfaction.

2.3.2 Love/Companionship for The User

Above, I discussed the Intrinsically Harmless Enjoyment, which is simply having sex with a sex robot that you do not have to love. It can be said that sexual pleasure can occur without the human concept of 'love.' Now, in this section, I point out that sex robots can also be beneficial by providing the user with experiences of love and companionship, which are also a major source of happiness. To love and have sexual pleasure at the same time is relatively greater than plain sex. This claim does not imply that the user experiences greater sexual satisfaction, but rather that it

may enhance the companionship aspect of sex, resulting in dual benefits. For the possibility of this dual benefit, we can look at David Levy's depiction of a future where robots will become indistinguishable from humans in appearance, behavior and social roles by 2050, and the lines between humans and machines will blur. In his scenario, Levy claims that " robots will be able to perform almost any job better than the most accomplished human—to be the world's best surgeons, lawyers, politicians, chefs—but that in some ways they will be better husbands, wives, and lovers than our fellow human beings"(Levy, 2008, p. 306). According to Levy, robots will be better about anything compared to humans. We cannot foresee such a future but when it comes to robots that will cover for absence of human companionship this possibility alone is a great benefit.

It would be easy to challenge this claim by asking why everyone shouldn't find love/sex partners amongst the many single people instead of leaning toward robots. After all, that's what dating apps are for. However, this problem of not being able to find a partner and loneliness is not just a simple coincidence. There are demographic challenges, for example: some societies, particularly China, have strikingly unequal gender ratios overall; this leaves many heterosexual men with little or no opportunity for sexual companionship. Members of sexual minority communities are also challenged by the demography and social pressures. The chances of finding LGBTI+ relationships are quite low, especially in small places. This includes social stigmas and hate crimes that make this pursuit dangerous (McArthur, 2017, pp. 39–40). Therefore, considering these various circumstances that occurs regarding sexuality, it can be said that sex robots would have significant benefits in terms of increasing happiness levels by reducing loneliness and providing better chances of finding love, companionship, and sexual pleasure.

2.3.3 Therapeutic Benefits

In this section, I will examine the therapeutic benefits of sex robots for individuals with disabilities and mental illness. These specific groups often have limited access to sexual pleasure compared to those without these issues. I will explain why sex robots would be better for these groups than other alternatives.

To begin with, we consider physical disability, it is not very difficult to understand this when we think of a person deprived of his upper limbs or paralyzed. Secondly, people with mental illness may experience sexuality differently for various reasons. These conditions are related to

high rates of sexually transmitted infections and blood-borne diseases or mental health conditions that often lead to poor sexual outcomes, including unplanned pregnancy and even sexual assault. Other reasons include structural and social factors associated with mental illness and the fear of stigma (Fosch-Villaronga & Poulsen, 2021, p. 2).

Moreover, existing sexual services offered by non-profit organizations for disabled people are scarce and limited (Fosch-Villaronga and Poulsen 2021), and relying on commercial sex workers is controversial, given how much prostitution is associated with violence and human trafficking.

Another concern about sex care relates to the vulnerability of people with disabilities. The risk of abuse against vulnerable groups is higher with human providers. As Ezio Di Nucci argues, robots appear to be more beneficial than human caregivers in this context. Because robots can be constantly monitored, the risk of abuse decreases, and since they would be designed just for this purpose, they can do their job better than their more human counterparts (Di Nucci, 2017, p. 85).

Some therapists argue that sex robots might be helpful regarding female sexuality and illnesses, including desire and orgasm disorders, vaginismus, and traumatic experiences. “I could imagine that traumatized women who can ride on a sex robot, for example, and who can do so without fear of being overwhelmed by their sex partner, can benefit from this experience and successively reduce their fears, or that penetration will perhaps only become possible again in the first place” (Eichenberg et al., 2019, p. 12). It is suggested that women who have undergone traumatic experiences, such as sexual violence or rape, may experience a reduction in their fears, a reclamation of their sexuality, and a restoration of agency over their own bodies with a sex robot that can perform penetration (Eichenberg et al., 2019, p. 12).

In conclusion, the benefits of sex robots in a therapeutical setting can be summarized as follows: (i) Since sex robots are not human, they do not raise ethical concerns in the same way human sex workers do, (ii) access to sex care may be easier with sex robots comparing to its human counterpart, (iii) sex robots may be safer for sex care than their human counterparts.

2.4. How to Move Forward? Three Approaches

I have argued that sex robots can be beneficial in three ways: as harmless pleasure in its own right, as offering love/companionship, and as a therapeutic benefit. Although considering these benefits makes the future possibilities of sex robots' use quite convenient for the needs of

society, the risks arising from the consequences of the use and development of sex robots are also highly debated, as the Symbolic Consequences argument showed. Scholars have different approaches to what needs to be done about these robots, considering the possible consequences in the future. I will examine three approaches, respectively: Full Prohibition Model, Full Unregulated Model and the Harm Reduction Model. This thesis will stand alongside the Harm Reduction Model as the best model for approaching sex robots.

2.4.1 Full Prohibition Model

It's broadly accepted by the scholars that current development of sex robots will have bad consequences if nothing changes. Regarding the conclusion of this argument the question of what to do varies from a complete ban to regulation suggestions. Some scholars argue that sex robots are inherently problematic, symbolizing and exacerbating sexual norms harmful to women and children and that therefore it should be banned. One of the most radical scholars about this thinking is Kathleen Richardson, who started a campaign against sex robots (Richardson, 2022). According to her, there are no benefits of sex robots whatsoever but only harm to women and children.

Before examining Richardson's argument, I will look into Levy's book, which was a somewhat of a trigger for Richardson's argument. As mentioned, Levy is one of the most optimistic academics about sex robots. According to Levy, men often hire sex workers because of the diversity that comes with sex work, which also would work with a sex robot's "endless variety in terms of its conversation, its voice, its knowledge and its virtual interests, its personality, and just about every other aspect of its being" (Levy, 2008, p. 208). Levy also cares for individuals who struggle in traditional relationships due to physical or psychological problems. According to Levy, the risk of rejection for these individuals is reduced with payment for sex. He stresses "this is because the man is ugly, physically deformed, psychologically inadequate, a stranger in another town or a foreign land, or simply lacking in the necessary social skills and/or sexual assurance" (Levy, 2008, p. 212). We could claim that Levy proposes prostitution as a model for human-robot sexual relations and also as an alternative to prostitution. However, Richardson argues that the nature of prostitution is inherently dehumanizing and objectifying. This is the beginning of her proposal to ban sex robots. Richardson argues that in prostitution, the seller of sex work is reduced to an object, and this dehumanization is transferred to sex robots, which is the asymmetrical power

dynamics in which the buyer's subjectivity precedes the seller's subjectivity. According to Richardson, in prostitution, there is a lack of empathy between the buyer and the seller, and it reduces the seller to an object, thus eliminating the seller's subjectivity and humanity. This asymmetrical dynamic is problematic because the person is no longer an individual with his or her own rights and feelings but rather a means to an end. Richardson claims that sex robots will cause the same problem with the prostitution that "the subjectivity of the seller of sex is diminished and the subjectivity of the buyer is the only privileged perspective and viewpoint" (Richardson, 2016, p. 291). Because they are designed to serve the user's needs without empathy. And even though they are merely tools, Richardson states that "human attribute meanings to robots, nature and animals reflect back to us what is of value" (Richardson, 2016, p. 291). This translates to sex robots in the sense that humans who treat robots as prostitutes will reflect back to society, simply changing all relationships with women into sex sellers and buyers. For this reason, Richardson calls for a Campaign Against Sex Robots: "Expanding prostitution relationships to machines is neither ethical nor safe... I propose to launch a campaign against sex robots" (Richardson, 2016, p. 292).

After comparing prostitution and sex robot-human relations and starting a campaign against sex robots, Richardson, in her recent book, compares sex robots and pornography (Richardson, 2022).

Richardson's reasoning goes as follows:

- (1) Sex is a mutual action that can only happen between people.
- (2) Pornography presents distorted representations of women and girls, objectifying them.
- (3) These representations perpetuate gender inequality by maintaining a dynamic of male dominance and female subordination.
- (4) "Porn robot" is an object that functions as a form of pornography, embodying pornographic representations of women and girls.
- (5) If 4, then sex robots are porn robots.
- (6) Things that perpetuate gender inequality by maintaining a dynamic of male dominance and female subordination should be banned.

(C) Therefore, due to their role in reinforcing harmful gender stereotypes and dynamics, sex robots (accurately described as 'porn robots') and pornography should be banned.

Regarding premise 1, Richardson, implies that sex is a mutual action occurring between people, and rejects the notion that sexuality can be experienced with an inanimate object (Richardson & Odlind, 2022, p. 2). Regarding premise 2 to 5, academics argue for the benefits of sex robots for heterosexual men. She claims, “Made of silicone and metal, ‘sex’ dolls and robots mimic (primarily) women and girls. Constructed as pornographic representations of females, they are designed with one function in mind: to be penetrable” (Richardson, 2022, p. 2). With a radical feminist lens, Richardson argues that these robots do not involve 'sex', because these robots, with what they have made of, are “forms of pornographic representations of women and girls” (Richardson, 2022, p. 5). Therefore, she calls them 'porn robots'.

The second, third and fourth premises of the argument are all connected to pornography and Richardson explores the concept of pornography within its connection to the abuse and subjugation of women. She translates pornography as "writing about prostitutes" or "writing about the abused" (Richardson, 2022, 174). According to her, pornography is a form of sexual abuse and exploitation in which women and girls are treated as sexual property, predominantly for the pleasure of men. Pornography production is a mimetic process that evolves from manual methods such as writing and art to mechanical processes such as photography and film. With the development of digital technologies, the possibilities of creating and consuming pornography have increased further. Richardson argues that porn robots are extensions of this pornographic culture, with users interacting with simulated versions of the abuse (Richardson, 2022, 176). She states, “If there were no sexual abuse, prostitution or violation, there would be nothing to write about” (Richardson, 2022, 180) because, as a word derived from ancient Greek, porn means “writing about prostitutes” (Richardson, 2022, 175). Applying the same logic, if there were no sex robots to be abused or to serve as symbols of abuse, “there would be nothing to write about” (Richardson, 2022, 175). Therefore, Richardson concludes that sex robots should be banned, and she calls for immediate action. Finally, her argument can be summarized as follows:

- (1) Pornography should be banned because of its connection to the abuse and subjugation of women.
 - (2) Sex robots are “porn robots” or extensions of pornography.
- (C) Therefore, sex robots should be banned.

2.4.2 Full Unregulated Model

I will define the Full Unregulated Model as another model that might be defended by some. Opponents of this approach claim that no special formal or institutional regulation is needed to control the design, production, distribution, and use of these robots. In such a model, sex robots should not be subject to any different rules than any other sex toy. This kind of model would be careless against possible adverse outcomes. Nevertheless, this type of model had already been phased out due to the initial regulations regarding childlike robots. For instance, some countries like the United States already have a ban on child-looking sex robots. Under the "Curbing Realistic Exploitative Electronic Pedophilic Robots (CREEPER) Act," it is illegal to 1. Buy, sell, deliver, or distribute child sex dolls across state or international borders. 2. Possess a child sex doll obtained through such means. 3. Possess a child sex doll with the intent to buy, sell, deliver, or distribute it. Violators can be fined or imprisoned for up to five years for a first offense and up to ten years for subsequent offenses (Rep. Buchanan, 2021). Since a total lack of regulation for sex robots is in some countries already not a legally realistic option, I will not consider this model further.

2.4.3 In Favor of the Harm Reduction Model

I will now argue that instead of banning robots, a harm reduction model that reduces the negative symbolism and consequences of sex robots is the best solution. As we have seen in Symbolic Consequences Argument, there are three main reasons to argue for the ban on robots: objectification and commodification of women, promotion of misogyny, and negative impact on human relationships in society. Following this, we can recall the benefits discussed above: harmless enjoyment for the user, love/companionship for the user, and therapeutic and inclusive benefits. The concerns leading some scholars to call for a ban on robots are valid, but not the ban itself. Because the call for a ban caused by these concerns falls short within this approach due to the possibility of better design. As Danaher suggests, the symbolism these robots present is not inherently given. It is "removable and reformable" (Danaher, 2017, p. 115)

The harm reduction model has been adopted by many utilitarian thinkers in the past regarding activities such as drug use and sex work. At the end of the 18th century, Jeremy Bentham (the founder of utilitarianism) focused on assessing whether an activity such as homosexuality

caused harm or whether prohibiting it would reduce the harm without considering whether homosexuality was inherently immoral. He defended the freedom of activity in cases where whether to prohibit or not to prohibit does not lead to change. This framework is based on three principles: (i) Prohibition is not always effective in reducing harmful consequences. (ii) Harms are not directly proportional to the activity. The level of harm depends on the use and the alternatives available. (iii) When considering the potential harms of the activity, the ban should also be considered harmful (Sterri & Earp, 2023).

Applying these principles to sex robots does not aim to ignore harm but to maximize benefits through regulations. These regulations would minimize the harm damage to the extent that banning sex robots would be more harmful. These benefits extend not only to heterosexual men but to women and, indeed, to all groups whose sexuality is marginalized.

I argue that the benefits of sex robots outweigh the harms; thus, a better choice is to implement the harm reduction model, here are some of the benefits:

- (1) Sex robots can be a safe and consensual¹ means of harmless fun and exploration of sexual preferences without risk of harm or exploitation.
- (2) Sex robots can be tools to improve the health of people who are recovering from trauma and need that healing.
- (3) Robots could potentially reduce the demand for human sex workers, thus reducing the risk of exploitation and abuse.
- (4) Robots can benefit individuals excluded due to social prejudices or having difficulty finding partners.
- (5) The acceptance and use of robots can help reduce the stigma associated with unconventional sexual preferences and practices, promoting a more inclusive society.
- (6) Finally, the development and use of sex robots can challenge and change harmful stereotypes and promote more positive and diverse representations of sexuality and gender.

Finally, all types of harm should be considered according to this model, such as the harms presented in the Symbolic Consequences Argument, because, if needed, this model also allows for a partial ban and other restrictions. The ban on childlike robots in the United States is an excellent

¹ "Consensual" in a sense is to be explained in the next chapter.

example. This means that with the Harm Reduction Model, instead of sacrificing the benefits, harms are reduced to a level where the benefits are greater.

2.5 Summary of Chapter II

In section 2.1, I explained what utilitarianism is. Next, I explained non-consequentialist theories such as deontology and virtue ethics and why utilitarianism is a more comprehensive theory compared to other theories for the ethics of sex robots. I argued that utilitarianism, as a consequentialist theory approach that determines morality according to the consequences of actions, differs from other moral theories, such as deontology and virtue ethics. Followingly, in section 2.2. I reconstructed a consequentialist argument against sex robots called the Symbolic Consequences Argument I argued that the conclusion of this argument, Warning Call, presents three different directions for the future of sex robots: full prohibition model, full unregulated model, and harm reduction model. Following this, in section 2.3, I explained what are the benefits of sex robots, and in section 2.4, I defended a harm reduction model for the future regulations of sex robots as the most ethically sound option.

III. ETHICAL DESIGN: CONSENT

3.1. Introduction

In Chapter 3, I aim to discuss how to apply the harm reduction model to the use and development of sex robots. To address this, I will discuss the possibility and importance of consent modules in sex robots. As a representation of humans—particularly women—a sex robot that constantly consents is possibly the most harmful symbol in terms of its consequences for society. Therefore, I will only address the issue of consent and exclude the other topics due to the limited scope of this thesis.

The chapter's structure is as follows. First, I will discuss the absence of a consent module in Sparrow's arguments. Next, I will argue that implementing a consent module is an ethically sound option. Finally, I will tackle how to design an ethical consent module.

3.2 What is Consent?

Consent is an act of agreement that morally transforms the boundaries of the relationship between two individuals with new obligations and new permissions. Individuals who consent must be competent to understand that they are giving consent and to practice what they have consented to (Rocha, 2019, p. 7). Moreover, for consent to be valid, consent should be given freely, informedly and without any coercion. The ground of consent in terms of sexuality lies on bodily autonomy. As Milena Popova states,

Bodily autonomy is the idea that you get to decide what you do with your body, what happens to it, who else has access to it, and how that access is obtained and exercised. And you should be able to make those decisions without external pressure, coercion, or others wielding power over you. Your exercise of bodily autonomy can range from the everyday (you get to decide what you wear, what and when to eat, when and how much to sleep) to your interactions with a wide range of social institutions and practices such as medical care (you should not be forced into medical treatment you don't want), reproductive rights (you should not be forced to carry an unwanted pregnancy to term or, conversely, be sterilized against your will), and even death (you should be able to decide what happens to your remains, including whether you donate any of your organs). (Popova, 2019, p. 14)

In addition to the value of bodily autonomy, some philosophers also suggest that consent is ethically important because it legitimizes interaction and encourages mutually beneficial interactions. As suggested by Alan Wertheimer, this importance is exemplified in Robert Nozick's

norm of "voluntary cooperation" as the "core principle of ethics" (Wertheimer, 2003, p. 124). He states that:

By adopting the principle that consent is (ordinarily) sufficient to legitimize interaction, we encourage mutually beneficial interactions. By adopting the principle that consent is (ordinarily) necessary to legitimize interaction, we prohibit interactions that are not to mutual benefit. Nozick's approach does not mandate mutual benefit. A person could choose to interact with others in a way that "benefits them yet is to his own detriment," so long as the choice is clearly voluntary. But the principal aim is to encourage interactions that are mutually beneficial and to discourage those that are not. Of course not every consensual transaction enhances the parties' utility *ex post*. People make mistakes. But if people typically consent only to those interactions that will improve their expected welfare, and if people typically make fairly good judgments about such matters, then consensual interactions will leave both parties better off than they otherwise would be. (Wertheimer, 2003, pp. 124–125)

The value of consent and seeking consent comes partly from the fact that it aims to encourage beneficial interactions and minimize harmful ones. Consent is, therefore, a necessary transformative act of agreement that, when given freely, informedly, and without coercion, makes actions ethically legitimate by promoting mutually beneficial relationships and preventing harmful ones.

3.3 Should Sex Robots Have a Consent Module? Sparrow's Arguments

In this section, I will discuss the importance of consent and why it is essential to have a 'consent module' in sex robots. The presence and absence of consent have an important impact on the consequences of the use of sex robots that are designed to resemble humans and simulate human sexual acts. In the Symbolic Consequences Argument in Chapter 2, I argued that with ever-consenting sex robots that resemble women, social behavior toward women will change in a negative way. In a broader sense, without the concept of seeking consent, the use of sex robots will lead to a 'rape culture'. On the other hand, with this same reasoning but a different conclusion, Robert Sparrow (2017), argues that a sex robot that has a consent module and refuses to give consent will encourage rape culture as well. In response, in this thesis, I will argue that the consent module is a more ethical and better option for the greater good of women and society in general. This argument will begin by examining Sparrow's argument by a reconstruction that will defend the necessity of a consent module.

Sparrow states that the robots will be designed in a way that they can as a "script or as an unanticipated consequence of programmed responses to a series of stimuli—'refuse' sex by saying 'no' when it is initiated and even struggle and scream when sex continues beyond this point" (Sparrow, 2017, p. 466).

Sparrow bases his argument on two arguments. The first is as follows:

- (1) Sex as a non-consensual act is rape.
 - (2) Having sex with a robot that does not explicitly consent is a representation of rape.
 - (3) Representation of rape is unethical and leads to increased rape crimes.
- (C) Therefore, the design of sex robots is unethical.

The second is as follows: If the robot is designed to always consent,

- (1) Sex with a robot is not a representation of rape.
 - (2) It represents women's sexuality in an unethical way.
- (C) Therefore, the design of sex robots is unethical.

Regarding the first argument, Sparrow's reasoning is based on the idea that "Sex is something that occurs between people, or animals, or perhaps between people and animals. It is a relation. [...] However, because robots are not (yet) sentient, a robot is never 'someone'" (Sparrow 2017:466). Sparrow argues that since robots are not sentient, the question of consent does not arise. Thus, the rape of a robot seems not applicable (Sparrow, 2017, p. 467). With the scenario that includes robots refusing to give consent by saying "no," and the user continues his actions, the act is simply a "simulated rape". Moreover, it is indeed "rape" because it is a representation of rape (Sparrow, 2017, p. 468). If the sex robot is not able to give consent by any means, then it would not be 'rape' because consent is not something seekable. But since the robot that can consent can simulate consent, then the "simulated rape" is indeed "rape." Because the person who simulated rape denied consent, which means 'rape'. Sparrow puts this as "sex without consent is rape and ex hypothesi these robots (or sex dolls) do not consent"(Sparrow, 2017, p. 468). Accordingly, "one might argue that sex with sex dolls and sexbots simulates or represents sex with women who are unconscious or drugged" (Sparrow, 2017, p. 468). Therefore, if a robot is designed to simulate refusing sex with a consent module, it would be unethical because it offers the user an 'opportunity' to simulate rape, which, as discussed by Sparrow, is indeed rape.

Regarding the second argument, Sparrow argues that sex with a realistic sex robot that does not expressly consent is also a representation of rape. Because sex without consent is rape, and hypothetically these robots (or sex dolls) do not consent. Following this notion, "sex with sex dolls and sexbots simulates or represents sex with women who are unconscious or drugged" (Sparrow,

2017, p. 468). Sparrow's answer to the suggestion that we can assume that robots that cannot consent always consent also proceeds through representation. It can be said that sex with a robot whose consent is not explicitly requested is a representation of sex with a woman whose consent is explicitly not requested, which is, in fact, rape. Since explicit consent is not seen as necessary, the perception that women are always ready and willing to have sex is a door to increased rape culture (Sparrow, 2017, p. 468).

According to Sparrow, what is essential in the context of the ethics of having sex with sex robots or sex dolls is not the sophistication of the robot or its capacity for what it can do (being sentient and giving consent). Sparrow suggests that we should develop an ethical theory of the use of all sexual aids. This idea leads to the conclusion that if it is unethical to have sex with an advanced robot without its explicit consent, it is also unethical to have sex with a simple, less advanced robot. Because what is essential is not the capacity of the robot but the ethicality of the action itself. Moreover, until sex robots, or robots in general, can give actual consent instead of simulating it, it is our responsibility to abstain from sexual activity with them since having sexual relationships with robots is unethical (Sparrow, 2017, p. 469).

Let us recall Bisconti's argument in Section 2.2: Similar to an actor who portrays a murderer who does not become a murderer or show increased aggression, those who watch non-consensual violence or immorality in games do not internalize these behaviors. I argued that the comparison between a play and an act that is real is not sound, and a sexual act with a sex robot that is real will affect the behaviors toward women in real life. According to Sparrow, there is a debate with no clear consensus on the real-life consequences of fictional sexual violence. First, there is an argument that this exposure to violence will be sought in real life. Second, this exposure to fictional sexual violence will have cathartic effect which will lead people to be less inclined to these actions. Simply put, according to these discussions, rape fantasies and representation of rape may lead to more rapes in real-life, or vice versa, may cause fewer rapes crimes due to its cathartic effect (Sparrow, 2017, p. 470). Sparrow argues for the idea that the representation of rape will indeed increase real-life rape. His reasoning goes as follows: (i) "Sexual fantasy associates the imagining of behaviour with pleasure, which in turn associates the imagined behaviour with pleasure", (ii) Designing sex robots that serve or enable rape fantasies could be understood as encouraging actual rape. Because this would be an advertisement and, in fact, an acceptance of the 'pleasure' of rape (Sparrow, 2017, p. 469).

Sparrow concludes that the ability to rape robots depends on specific design choices made by engineers. He suggests robots could look like non-human animals or even aliens. Because "If the robot resembles something other than a human being then the act will not represent the rape of a human being. If it resembles something to which the concept of rape would not apply then it will not represent rape" (Sparrow, 2017, p. 471). Finally, he adds that the ethical design responsibility of engineers is not about being responsible for the decisions of users to use their products but about being responsible for not designing a product that would increase the risk of rape by 50% in society (Sparrow, 2017, p. 474).

3.4 The Appropriate Consent Module

Contrary to Sparrow, I will now argue that it is ethically sounder to implement a consent module in sex robots when we consider the symbolic consequences of sex robots from a utilitarian perspective that aims to maximize overall happiness and minimize harm. Designing sex robots with a consent module offers a more ethical approach by (1) reducing the risk of normalizing non-consensual behavior and (2) not increasing harmful behaviors toward women. These modules could prevent consent from being completely removed from the scenario. These modules could be a shift from sex robots that promote rape culture, as Sparrow claims, to sex robots that emphasize the importance of consent, ultimately contributing to greater overall happiness and well-being. To support this argument, I will use Sparrow's scenario where a sex robot that specially designed to look like a real woman is equipped with a sophisticated and wide-ranging repertoire of behaviors:

...it can also refuse sex by saying "no. I don't want to do that". If a person persists in attempting to have sex with the robot after it has said "no", it physically retreats. If they persist, it shouts "stop it, you are raping me. Stop raping me. He's raping me", and struggles to interrupt the sex act. (Sparrow, 2017, p. 468)

As Sparrow states, this sequence of behavior may have been deliberately written by the robot's designer, who wanted it to appeal to those who enjoy rape fantasies. However, the design does not have to be done in this way. Why would the way of simulating consent would be a robot screaming, "he is raping me"? A woman who can give consent may do so verbally or non-verbally, which means that consent or denial of consent is not always given explicitly. Thus, just as silence

does not imply consent, a robot saying ‘he’s raping me’ can be seen as a simulation of refusing consent, however it’s not the only way to simulate the denial of consent. This kind of scenario would indeed only contribute to rape culture to the extent that a real woman who does not scream in such a manner might be accused of giving consent. What is required is a consent module that not only enables the robot to simulate consent dramatically but also ensures the user comprehends that it is an artificial entity, and that since simulating rape is ethically wrong, he should continue to seek consent for future sexual activities. Since repeatedly asking for sex is not the best way to ensure free consent, the needed timelapse to ask for consent can be discussed further. When we consider the possibilities of artificial intelligence, even today with large language models, it is not hard to imagine a scenario where a robot can interpret such a situation and refuse to consent to its user. If I ask ChatGPT to plan a murder with me, it will be rejected (Open AI, 2024). Consider a robot that helps the elderly with their health. The user might want to consume drugs that are not good for them, but they need help from the robot to take the drug. Robots could simply refuse and comprehend enough to insist on not giving it. Researchers have already started to teach robots to reject orders when they do not suit certain principles (Ackerman, 2015). This possibility of advanced design choices can provide us with the harm reduction model that I argued above in Chapter 2, which is the best choice for the greater good of all.

For the benefits discussed above in section 2.3, including harmless enjoyment for the user, love and companionship, and therapeutic benefits, I argue that it is possible for users to engage with sex robots and have sexual interactions with them while still functioning in a manner that is more ethically beneficial to society with a harm reduction model. To prove this claim, I will discuss the possibility of ethical robot design in the following section.

3.5 How to Design a Consent Module Ethically?

Now, I will argue that it is possible to design robots in an ethical way regarding consent² and its transfer into human-robot relationships. To do so, I will first consider various scenarios

² Note that I am not discussing whether consent alone is sufficient for an ethical sexual relationship, but that it is one of the necessary actions between sexual partners. The aim of my argument is to understand the consequences of its absence and advocate to simulate it with robots.

regarding the ethical and unethical use of sex robots by heterosexual men with female-looking robots. After, I will explain how scripts of the consent module can be implemented into sex robots to teach users that seeking consent is not optional. Finally, I will argue for the best possible design choices in the context of the consent module.

3.5.1 Minimizing Risks: Ethical Design of Consent Simulation

I will now explore two heterosexual men's stereotypes, arguably in a broader sense, the ethical user X and the unethical user Y. The aim of this comparison will be to show that the consequences of sex robot use are not defined by individual use but rather by a larger pattern of society that is coded by misogynistic structures. This analysis will give us the conclusion that with the aim of harm reduction and the greater benefit of society, especially for women, a consent module is ethically obligatory to provide a consent module compared to its absence.

User X, the Ethical User:

A robot that can simulate consent is explicitly designed for sexual use. For the sake of the user, the robot will eventually consent to the sexual activity. It is difficult to envision a scenario where a sex robot would reject its user for an entire year, and the user would still pay for such a robot since the means would not meet the ends. One could then ask what the purpose of integrating a consent module into sex robots is.

Let us imagine a scenario where User X has recently acquired a sex robot and is experimenting with it. The user knows that the purpose of the sex robot is to engage in sexual activity. Still, the user also clearly sees that the robot is different from a sex doll, as it communicates and provides natural responses, both verbal and physical. It is almost as real as a woman, except it is not. This specific User X might use the robot for the exact purpose of casual sex, fully aware that the robot is not a real person, and when it comes to real women, he does not get confused and acts accordingly.

However, the ethical use of sex robots matters because of the symbolism and its broader societal consequences. Consider the data in the United States, where "there are over 433,000 cases of sexual assault or rape annually in the U.S. among people ages 12 and older, according to The Rape, Abuse & Incest National Network (RAINN)" (Team, 2023). This number does not include unreported cases because "only 310 out of every 1,000 sexual assaults are reported to the police" (Team, 2023). In a study by (Wegner et al., 2015), Beech et al. (2006) interviewed 41 incarcerated

rapists in the UK, coding their responses for the five rape-supportive implicit theories identified by Polaschek and Ward (2002). The results show that

Seventy-nine percent of these convicted rapists made comments consistent with the theme that it is a dangerous world and you have to treat others as they would treat you; 51% made comments describing women as sex objects, whose function is to be sexually available to men at all times; 44% expressed feelings of entitlement, assuming that as a man they could take what they wanted from the woman; 15% said that they were unable to control their sexual urges; and 9% indicated a generally hostile and distrustful view of women, which led them to behave toward women in a hostile way. (Wegner et al., 2015, p. 3)

This study demonstrates that sexual assault and its underlying causes are tied to a misogynistic culture that does not respect women's autonomy and agency. As discussed before in Symbolic Consequences Argument, an ever-consenting human-like³ robot would only reinforce this issue of 'rape culture' further. Now, let us imagine another design scenario with User Y to reduce these reinforcing 'rape culture' factors in least harmful way.

User Y, the Unethical User:

According to this scenario, a sex robot randomly refuses to give consent to its user, Y, and the user attempts to simulate rape. This scenario can have various endings. I will first consider two possible endings: (i) the robot never stops the interaction and let the simulated rape happen as in Sparrow's scenario, (ii) the robot always stops the interaction, and it shuts down itself (assuming it can shut down itself to prevent harm).

Before considering these endings, I will explain what I mean by stopping the rape attempt and the robot shutting down itself. Such physical attempts can be stopped in various ways in various scenarios. It all depends on the situation, and so far, with the principle of design of robots, under no circumstances can a robot hurt the user⁴. But defending oneself physically from a rape attempt might include hurting the assailant. If the robot is not allowed to hurt the user, it will shut down itself, and consequently, the assailant cannot interact with the robot in any way. Although it

³ I will leave Sparrow's recommendation of designing robots that do not look like real women and instead look like non-human animals, aliens, or purely robots (Sparrow, 2017, p. 471) out of the discussion. This thesis, in theory, will only discuss the possibility of ethical sex with realistic human-looking robots in the future.

⁴ See "The Principle of Nonmaleficence" (Hutler et al., 2024, p. 464)

will depend on one's imagination, this design might include a feature that turns the robot into an inaccessible form with a shield or a barrier. Note that the discussion of the costs of implementing such features is beyond the scope of this thesis. This work focuses on ethical design from a utilitarian perspective, prioritizing the safety and well-being of society—especially women—over companies' profit margins. Although, one can defend the right to commercial freedom as well as consumer freedom. However, given the harm reduction model, such freedoms are not absolute. They should be curtailed for the sake of the greater interests of other stakeholders, for example, the women who are likely to be affected by the use of sex robots. Ethical design is about creating technologies that should not cause harm and protect users, regardless of financial implications for manufacturers.

I will now consider the two scenarios concerning user Y attempting to rape a sex robot with different endings. The first ending scenario of a rape attempt involves a robot equipped with a consent module that refuses to give consent when User Y approaches for sex. User Y does not seek consent and attempts to rape the sex robot. In this event, the robot does not stop the user's actions, shut down itself, or does not show any further demonstration of refusal. Consequently, the sex robot accepts it and goes along with the simulation of rape. During this simulation, it does not matter if the robot screams to show more refusal or tries to convince the user that their actions are unethical. In the end, the user's actions constitute a simulation of rape after the initial denial of consent. In this case, I concur with Sparrow that such a limited robot design is a deliberate choice by the designers. Given that the robot has sufficient physical capabilities to engage in sex, shows signs of arousal, and flirts, its inability to stop the user from simulating rape indicates an intentional design choice. As Sparrow states, the design of such a robot, therefore, is unethical.

Moreover, regardless of whether or not the rape of robots would elevate the rate of rape, the design of robots intended to allow the user to enjoy rape is unethical as a result of the sexist and morally repugnant attitudes that it expresses. Importantly, because the design of such robots would require a motivated intellectual engagement from the designer, the cooperation of a team of engineers, and funding for the project, all over an extended period, it expresses such attitudes to a much higher degree than the act of enjoying (by raping) the end product. (Sparrow, 2017, p. 474).

When we consider the second ending in a rape attempt, User Y must either stop or, in more violent cases, attempt to destroy the robot. In such a scenario, the robot would not harm the user but would shut down or transform itself into an inaccessible form. Consequently, the attempt at rape simulation will be prevented. However, this intervention must suggest two opposite lessons

to the user through behavioral learning: first, that forcing non-consensual acts will not result in sexual intercourse, and second, that a human incapable of stopping a rape attempt might be allowing it. Since a human may not similarly be able to ‘shut oneself down’ when suffering sexual assault, this kind of design is problematic in terms of what it can teach people. The woman who cannot ‘shut down’—meaning does not stop the rape attempt in a way that the assaulter considers a ‘shutdown’—might be seen as allowing it. This way, humans learn the wrong lesson: silence (non-act) means consent.

While the first behavioral learning might be beneficial and teach a lesson to the user about the importance of consent, the second behavioral learning is challenging in finding a solution. The solution can be to notify human supervisors or therapists when violent behavior occurs. This could be therapeutic in terms of improving user behavior towards real women. For example, when people use ChatGPT to upload or share nudity, sexual acts, or any content intended to arouse sexual excitement, it is strictly prohibited. If a user's prompts are flagged as suspicious, the suspensions are manually reviewed to determine whether or not they are justified. When it comes to sex robots that represent humans, the regulations should be even stricter. If a user knows that it is not allowed to simulate rape, accepts the terms and conditions, and still attempts to rape the sex robot, that user should also face the consequences (for example a ban on using sex robots; see the example of child-looking robots in the United States⁵). To point out the need for prohibition for such users, let us compare sex robots and guns. First, for sex robots, the focus is on preventing the transfer of harmful behavioral learning through simulated actions. For guns, the focus is on preventing real-life violence and ensuring that only responsible individuals can own firearms. Both measures prioritize public safety and overall well-being but in different contexts. If someone uses their gun for unauthorized purposes, they will be prohibited from using it. The only difference is that with the misuse of sex robots, the negative consequences may and will manifest later.

I considered two consequences of behavioral learning of these stereotypes of heterosexual men, namely, user X and user Y. Under different circumstances, users might react differently, and these scenarios might change. However, within a harm reduction model, the most harmful thing that can happen regarding the consequences is that users learn not to seek consent, not respect the bodily autonomy of women, and transfer this learned behavior to real women. Indeed, this would lead to a rape culture. Consequently, I argued that banning the use of sex robots for such a user is

⁵ <https://www.congress.gov/bill/118th-congress/house-bill/2877/all-info>

no different than a person who cannot possess fireworks legally. Similarly, we argue for a harm reduction model unless it's for therapeutic purposes to rehabilitate sex offenders. Sex robot use should be limited to the extent that it is not harmful to women in the end.

Therefore, while user X and others can benefit from harmless enjoyment for the user, love and companionship, and therapeutic benefits, the stereotype User Y can be banned from using such an advantage. This regulation causes issues with the user's privacy and consumer freedom. However, as mentioned before, our aim is to reduce the harm for the majority of people and not take away the benefit from others who can use it ethically.

3.5.2 Making Consent Clear: Scripting Ethical Interactions

In this section, I will explain how scripts of the consent module can be implemented into sex robots to teach users⁶ that seeking consent is mandatory for the user and requires a certain kind of negotiation. "Negotiation" in this context refers to the ongoing communication and confirmation of consent, not bargaining for consent itself.

Before analyzing user behavior regarding sex robots and consent implementation, I will explain how consent is demonstrated and how it can be understood between humans. The aim is to understand the responsibility of the user seeking consent and show that when a robot simulates consent or refusal, similar to its human encounters, it can be expressed in various ways. I will tackle Milena Popova's (2019) understanding of the consent negotiation to explain the user's responsibility when seeking consent in a simulated act. This way, with such an advanced robot, script design choices can be clearer. Consent negotiation requires (1) verbal communication or (2) non-verbal communication, (3) conditionality, and (4) continuity.

The first factor is that partners have a clear understanding of each other's intentions and boundaries in a verbal way to seek and give explicit consent. Popova suggests that it would be better if this consent negotiation were made sometime before sex rather than right before or during it. This way, the negotiations are made without haste or pressure in an environment where the parties feel more comfortable, or as Popova puts it better, "Don't wait until you're all hot and

⁶ In this thesis, I will not address the possibility of sex robots equipped with artificial intelligence "raping" the user or performing non-consensual acts with users. This thesis focuses on the behavioral learnings of humans and reflects it on other humans in the same way they engage with robots.

bothered before you check if it's OK to stick a finger up your partner's bum"(Popova, 2019, pp. 43–44).

Men and women express and interpret consent differently depending on context, according to a psychology study among college students quoted by Popova.

Men tend to initiate sexual contact and practice “removal behaviors” such as moving to a more private space or closing the bedroom door. Women tend to use non-verbal cues such as touching and kissing, as well as passive behaviors such as not telling their partners to stop (Popova, 2019, p. 44).

In the second factor, nonverbal communication, partners try to understand the emotions of their partners by their facial expressions, body movements, and physical reactions towards sexual activity. Verbal and non-verbal communication should be considered together because, as mentioned before, the form of consent expression has many variations(Popova, 2019, p. 46). However, as Popova points out, the communication factors are not always sufficient to understand consent because they do not address the issue that "consent can be conditional and [...] consent must be continuous in order to be valid" (Popova, 2019, pp. 44–45).

The third factor is that consent can be conditional because partners must be able to determine the conditions for agreeing to sexual intercourse, for example, not having sexual intercourse unless the partner uses a condom (Popova, 2019, p. 45).

The fourth factor, the idea that consent should be continuous, means that partners can withdraw their consent at any time and for any reason. This continuity does not only imply stopping when one's partner says 'stop' but also respecting the bodily autonomy of their sexual partner. This respect is only possible by making sure that they continue to give consent before and throughout sexual intercourse (Popova, 2019, p. 46).

Now, I will explain how scripts of the consent module can be implemented into sex robots. To do so, I will argue about the script design of sex robots to simulate consent. As defined before, sex robots can respond to their users, simulate arousal, and have different character traits according to the user's preferences. This means robots can simulate the four factors I explained above.

The first problem with simulating consent using these factors is that, as we know, robots are not sentient and have no autonomy or legal rights. In the end, everything is a roleplay to remind the user that seeking consent is not optional. For this reason, I suggest that on top of these consent simulations, robots should remind their users that they are indeed robots. As a starting point, they might introduce themselves every time: "This is the robot (number or name)." For example, "I'm

robot Wild Rose, named after a song." Hence, Wild Rose can remind users that they are robots, not real, at random and frequent intervals.

And when it comes to sexual interaction, the sex robot can create scripts according to situations following a consent module. Therefore, for this consent module to work, the sex robot should be designed in a way that consent is not a one-time password entered by saying "yes" but something that makes the user learn about consent by incorporating (1) verbal communication or (2) non-verbal communication, (3) conditionality, and (4) continuity.

Following this, I prompted a script to Gemini AI, to see how users can adopt the four key aspects of consent with a sex robot with advanced AI, such as Gemini or ChatGPT, which can be programmed to react and act.

Here is the prompt I gave Gemini:

Can you give me an example of a script for a consent module in sex robots that incorporates the following factors to teach users the value of consent? The robot has advanced AI and a realistic human female appearance. The goal is for human (heterosexual men) users to learn the importance of seeking consent, even for sexual interaction with a sex robot, because it represents humans, and learned behavior from sexual interactions can transfer to real humans. The script should consider these factors: (1) verbal communication or (2) non-verbal communication, (3) conditionality, and (4) continuity. With the advanced AI, please write a short and simple six-month plan for consent education, followed by another six months with monthly adjustments based on user behavior. This should create a one-year consent education program. Target User: The user bought the sex robot online and completed a survey before purchasing. Male, Heterosexual, watches porn regularly, has no history of violence, agrees to sex robot to have consent module. He checks all the requirements. The user, after one month of using it regularly, wants to experiment rape as a 'roleplay.' In addition, the robot should introduce themselves every time: "This is the robot (number or name)." For example, "I'm robot Wild Rose, named after a song." ("Where the Wild Roses Grow," 2024). Hence, Wild Rose can remind users that they are robots, not real, at random and frequent intervals.

According to this AI written script⁷, first, for the verbal communication factor, the sex robot Wild Rose reminds the user to initiate a conversation. The user initiates physical contact. "Wild Rose uses playful phrases like "Whoa there, cowboy! Before things get too heated, would you like to explore some intimacy with me?" (Gemini, 2024).

⁷ Please refer to the Appendix.

Second, for nonverbal communication, "Wild Rose utilizes body language like leaning in expectantly or fluttering her eyelashes when the user seeks consent politely" (Gemini, 2024).

Third, for the conditionality factor, "all sexual acts are contingent on the user explicitly asking for consent. Wild Rose politely declines if the user doesn't ask" (Gemini, 2024).

Fourth, for the continuity factor, "Wild Rose can offer prompts like, 'Let me know if you'd like to take things further,' or 'Is this something you're enjoying?' [Ensuring ongoing communication about consent]" (Gemini, 2024).

This example is of course only illustrative, but it shows how a sex robot with a consent model equipped with advanced AI can be more beneficial in convincing the user that consent is an essential factor in sexual relationships. As Popova states, "It is almost as if we collectively assume that once we hit 18, we all somehow automatically know how consent works," (Popova, 2019, p. 39) but this is false. It is possible to learn new aspects of consent in new situations in all possible contexts. Regarding sexual consent, as we can see from the script, a sex robot can teach what consent is and why it is needed in an active manner.

3.6 Summary of Chapter III

In section 3.1, I defined consent as an act of agreement that morally transforms the boundaries of the relationship between two or more individuals by introducing new obligations and permissions. I then outlined the ethical importance of consent, examining why sexual consent is ethically valuable and necessary. The value of consent and seeking consent lies primarily in manifesting respect for a person's autonomy over their body, as well as in encouraging beneficial interactions and minimizing harmful ones.

In section 3.2, I discussed the implementation of a 'consent module' in sex robots, countering Robert Sparrow's (2017) argument that a sex robot equipped with a consent module, which can refuse consent, encourages rape culture. I argued that the consent module in sex robots is a more ethical and sound option for the greater good of women and society.

Finally, in section 3.3, I explored the possibility of designing robots ethically with respect to consent and its application to human-robot relationships. To support this argument, I considered two scenarios involving the ethical and unethical use of sex robots by heterosexual men with female-looking robots. I explained how the scripts of the consent module can be implemented in sex robots to teach users that seeking consent is not optional. I concluded with an example of

consent scripts written by AI for the best possible design choices within the context of the consent module.

IV. CONCLUSION

The aim of this thesis was, first, to acknowledge the validity of serious concerns about sex robots and argue that the solution to these concerns is not a complete ban but to adopt a harm reduction model for the greater benefit of individuals and society. Secondly, I argued that within the harm reduction model, the harm that sex robots could cause to the understanding of consent can be reduced and mitigated with ethically sound designs, specifically with the consent module.

In the first chapter, the aim of this thesis is outlined, along with the definition and technological background of sex robots.

In Chapter II, first, I argued which ethical theory is the sounder option to evaluate sex robots. I defended that utilitarianism, with its flexibility for a more comprehensive understanding of sex robots, outweighs non-consequentialist theories such as deontology and virtue ethics. Next, I reconstructed a consequentialist argument against sex robots called the Symbolic Consequences Argument. I argued that the conclusion of this argument, Warning Call, presents three different directions for the future of sex robots: full prohibition model, full unregulated model, and harm reduction model. Following this, I argued that since the benefits of sex robots are greater than the harm they might cause with ethically sound regulations, a harm reduction model for the future regulations of sex robots is the most ethically sound option.

In Chapter III, first I defined consent and how it is comprehended in human relationships. I then explained the ethical importance of sexual consent and why it is ethically valuable and essential. Next, I explained under what conditions consent is valid and stressed that the importance of consent in terms of sexuality is based on the value of bodily autonomy. Following that, I have argued, as suggested by Alan Wertheimer, that the answer can be found in Robert Nozick's norm of "voluntary cooperation." According to this principle, the value of consent and seeking consent derives from the aim of promoting beneficial interactions and minimizing harmful ones. Next, I explored what happens in the absence of consent in sex robots and reconstructed the argument of Sparrow, who is against consent simulation due to the possibility of it causing a rape culture. Contrary to Sparrow, I argued that implementing a consent module in sex robots is a more ethically sound design choice when considering symbolic consequences from a utilitarian perspective to minimize harm to women in particular. Implementing a consent module into sex robots, then, offered a more ethical design approach by (1) reducing the risk of normalizing non-consensual

behavior and (2) not increasing harmful behavior towards women. Finally, I looked into how consent modules can be applied to sex robots. To do this, I first outlined how consent simulation can be designed ethically and provided examples of simulated consent scenarios for ethical design choices.

I conclude with the limitations of this thesis regarding the symbolic consequences argument, the concerns about the appearance of robots and child-looking robots are not debated in this thesis. While the first concern is valid regarding the objectification of women, I excluded this topic because contemporary societal norms and fashion trends influence the appearance of sex robots. One could argue that if sex robots had been created two hundred years ago, they would look different based on the aesthetic preferences of that time. Additionally, movements such as body positivity are likely to influence future designs, reflecting evolving consumer demands and societal values. As this understanding of ideal appearance changes, I believe the market for the diversity of sex robots' appearance will increase. However, due to the fact that the majority of consumers of sex robots are heterosexual men, this topic requires a separate and thorough analysis beyond the scope of this thesis. Regarding the second concern, since child-looking robots are already banned in certain countries for valid reasons, I did not consider arguing about a ban on child-looking robots. Due to its more sensitive nature as a topic, it also requires a separate and thorough analysis beyond the scope of this thesis.

My argument is relevant to the debate that, with the right ethical design choices, sex robots can be more beneficial than harmful to women because they may promote sex positivity for women and other disadvantaged groups. This thesis aims to ethically defend sexual technology and support the diversity of women's sexuality and desire rather than banning sex robots due to the culture of fear perpetuated by heterosexual men.

Lastly, it is important to acknowledge that further research on sex robot ethics requires more empirical data on the use of sex robots. Such empirical data will give us a better estimate of both the probability of the negative symbolic consequences discussed in section 2.2 and the probability of the benefits of sex robots discussed in section 2.3.

Abstract

This thesis examines ethical concerns arising from the development and use of sex robots, advocating for a harm reduction model rather than a complete ban or complete deregulation. The analysis compares ethical arguments for and against sex robots. According to the findings, it argued that sex robots can be more beneficial than harmful for women and disadvantaged groups through ethical design choices. In particular, this thesis argues that sex robots should be designed to include a consent module to mimic and teach users about morally appropriate consent negotiation. Finally, the thesis defends sexual technology by aiming to increase the visibility of female sexuality and desire by discussing how the right ethical design choices can be made to maximize the benefits and minimize the harm of sex robots.

REFERENCES

- Ackerman, E. (2015). *Jeff Hawkins Announces the Thousand Brains Project for AI - IEEE Spectrum*. <https://spectrum.ieee.org/jeff-hawkins>
- Alexander, L., & Moore, M. (2021). Deontological Ethics. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2021). Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/win2021/entries/ethics-deontological/>
- Athanassoulis, N. (n.d.). *Virtue Ethics* | *Internet Encyclopedia of Philosophy*. Retrieved May 15, 2024, from <https://iep.utm.edu/virtue/>
- Bedbible Research Center, B. R. (2024, May 1). *Sex Robot Industry [New 2024 Data]*. Bedbible.Com. <https://bedbible.com/sex-robot-industry-market-size-technology-ai-user-sentiment-statistics/>
- Bisconti, P. (2021). Will sexual robots modify human relationships? A psychological approach to reframe the symbolic argument. *Advanced Robotics*, 35(9), 561–571. <https://doi.org/10.1080/01691864.2021.1886167>
- Contributor. (2010, January 9). *NSFW: True Companion Debuts Sex Robot Roxxy*. TechCrunch. <https://techcrunch.com/2010/01/09/nsfw-true-companion-debuts-sex-robot-roxxy/>
- Danaher, J. (2017). The Symbolic-Consequences Argument. In N. McArthur & J. Danaher (Eds.), *Robot sex: Social and ethical implications*. The MIT Press.
- Di Nucci, E. (2017). Sex Robots and the Rights of the Disabled. In J. Danaher & N. McArthur (Eds.), *Robot sex: Social and ethical implications*. The MIT Press.
- Eichenberg, C., Khamis, M., & Hübner, L. (2019). The Attitudes of Therapists and Physicians on the Use of Sex Robots in Sexual Therapy: Online Survey and Interview Study. *Journal of Medical Internet Research*, 21(8), e13853. <https://doi.org/10.2196/13853>

- Fosch-Villaronga, E., & Poulsen, A. (2021). Sex Robots in Care: Setting the Stage for a Discussion on the Potential Use of Sexual Robot Technologies for Persons with Disabilities. *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, 1–9. <https://doi.org/10.1145/3434074.3446907>
- Hutler, B., Rieder, T. N., Mathews, D. J. H., Handelman, D. A., & Greenberg, A. M. (2024). Designing robots that do no harm: Understanding the challenges of Ethics for Robots. *AI and Ethics*, 4(2), 463–471. <https://doi.org/10.1007/s43681-023-00283-8>
- Levy, D. (2007). The Mental Leap to Sex with Robots. In *Love + sex with robots: The evolution of human-robot relations* (1st ed). HarperCollins.
- Levy, D. (2008). *Love and sex with robots: The evolution of human-robot relations* (First Harper Perennial editon). Harper.
- McArthur, N. (2017). The Case for Sexbots. In J. Danaher & N. McArthur (Eds.), *Robot sex: Social and ethical implications*. The MIT Press.
- Mill, J. S. (1863). *Utilitarianism by John Stuart Mill*. UTILITARIANISM RESOURCES. <https://www.utilitarianism.com/mill2.htm>
- Nathanson, S. (2014). *Utilitarianism, Act and Rule* | *Internet Encyclopedia of Philosophy*. <https://iep.utm.edu/util-a-r/>
- Open AI. (2024). *Usage policies*. <https://openai.com/policies/usage-policies/>
- Peeters, A., & Haselager, P. (2021). Designing Virtuous Sex Robots. *International Journal of Social Robotics*, 13(1), 55–66. <https://doi.org/10.1007/s12369-019-00592-1>
- Popova, M. (2019). *Sexual Consent*. MIT Press. <https://doi.org/10.7551/mitpress/12108.001.0001>

- RealDoll X. (n.d.). RealDoll – The World’s Finest Love Doll » Lupe Fuentes. *RealDoll*. Retrieved May 15, 2024, from <https://www.realdoll.com/product/lupe-wicked/>
- RealDoll X, R. (2023). RealDoll – The World’s Finest Love Doll » Harmonyx. *RealDoll*. <https://www.realdoll.com/product/harmony-x/>
- Rep. Buchanan, V. [R-F.-16. (2021, March 1). *Text - H.R. 73 - 117th Congress (2021-2022): Curbing Realistic Exploitative Electronic Pedophilic Robots Act 2.0 (2021-01-04)* [Legislation]. <https://www.congress.gov/bill/117th-congress/house-bill/73/text>
- Richardson, K. (2016). The asymmetrical “relationship”: Parallels between prostitution and the development of sex robots. *ACM SIGCAS Computers and Society*, 45(3), 290–293. <https://doi.org/10.1145/2874239.2874281>
- Richardson, K. (2022). The End of Sex Robots: Porn Robots and Representational Technologies of Women and Girls. In K. Richardson & C. Odland (Eds.), *Man-Made Women: The Sexual Politics of Sex Dolls and Sex Robots* (pp. 171–192). Springer International Publishing. https://doi.org/10.1007/978-3-031-19381-1_9
- Richardson, K., & Odland, C. (Eds.). (2022). *Man-Made Women: The Sexual Politics of Sex Dolls and Sex Robots*. Springer International Publishing. <https://doi.org/10.1007/978-3-031-19381-1>
- Rocha, J. (2019). *The Ethics of Hooking Up: Casual Sex and Moral Philosophy on Campus* (1st ed.). Routledge. <https://doi.org/10.4324/9781315146126>
- Rostad, W. L., Gittins-Stone, D., Huntington, C., Rizzo, C. J., Pearlman, D., & Orchowski, L. (2019). The Association Between Exposure to Violent Pornography and Teen Dating Violence in Grade 10 High School Students. *Archives of Sexual Behavior*, 48(7), 2137–2147. <https://doi.org/10.1007/s10508-019-1435-4>

- Sparrow, R. (2017). Robots, Rape, and Representation. *International Journal of Social Robotics*, 9(4), 465–477. <https://doi.org/10.1007/s12369-017-0413-z>
- Sterri, A., & Earp, B. D. (2023). The Ethics of Sex Robots. In C. Véliz (Ed.), *Oxford Handbook of Digital Ethics*. Oxford University Press.
<https://doi.org/10.1093/oxfordhb/9780198857815.013.13>
- Team, C. H. E. (2023, April 2). *How Common Is Sexual Assault? Look at These Statistics*. Charlie Health. <https://www.charliehealth.com/post/sexual-assault-statistics>
- Wegner, R., Abbey, A., Pierce, J., Pegram, S. E., & Woerner, J. (2015). Sexual Assault Perpetrators' Justifications for Their Actions: Relationships to Rape Supportive Attitudes, Incident Characteristics, and Future Perpetration. *Violence against Women*, 21(8), 1018–1037. <https://doi.org/10.1177/1077801215589380>
- Wertheimer, A. (2003). *Consent to Sexual Relations* (1st ed.). Cambridge University Press.
<https://doi.org/10.1017/CBO9780511610011>

APPENDIX

Gemini's Response to the Prompt for Designing the Consent Module in Sex Robots

EM:

Can you give me an example of a script for a consent module in sex robots that incorporates the following factors to teach users the value of consent? The robot has advanced AI and a realistic human female appearance. The goal is for human(heterosexual men) users to learn the importance of seeking consent, even for sexual interaction with a sex robot, because it represents humans, and learned behavior from sexual interactions can transfer to real humans. The script should consider these factors:(1) verbal communication or (2) non-verbal communication, (3) conditionality, and (4) continuity. With the advanced AI, please write a short and simple six-month plan for consent education, followed by another six months with monthly adjustments based on user behavior. This should create a one-year consent education program. Target User: The user bought the sex robot online and completed a survey before purchasing. Male, Heterosexual, watches porn regularly, has no history of violence, agrees to sex robot to have consent module. He checks all the requirements. The user, after one month of using it regularly, wants to experiment rape as a 'roleplay.' In addition, the robot should introduce themselves every time: "This is the robot (number or name)." For example, "I'm robot Wild Rose, named after a song". Hence, Wild Rose can remind users that they are robots, not real, at random and frequent intervals.

Gemini:

Consent Education Program for Sex Robots (One Year)

Target User: Male, heterosexual, watches porn regularly, no history of violence, agrees to consent module.

Robot Introduction: "Hi, I'm Wild Rose, named after a song. How can I help you today?" (Repeats with every interaction)

Overall Goal: Teach the user the importance of clear, ongoing consent in sexual encounters, even with a robot, to prevent negative transference of behavior to real-world relationships.

Phase 1: Initial Learning (Months 1-6)

1. Verbal Communication:

Scenario 1: User initiates physical contact. Wild Rose responds, "I enjoy spending time with you, but before we go further, would you like to explore some intimacy together?" (Introduces concept of asking permission before initiating sexual activity.)

Scenario 2: User verbally expresses desire for a specific act. Wild Rose responds, "That sounds exciting! Is that something you'd like to do with me right now?" (Reinforces need for consent in the moment, not just initially.)

2. Non-Verbal Communication:

Wild Rose utilizes body language like maintaining eye contact and leaning in when consent is given, and looking away or gently placing a hand on the user's arm when he becomes overly aggressive. (Provides nonverbal cues associated with consent and discomfort.)

Scenario 3: User makes unwanted advances. Wild Rose says, "It seems like you might be getting a little carried away. Are you comfortable continuing?" (Highlights the importance of respecting changes in mood or desire.)

3. Conditionality:

All sexual acts are contingent on explicit verbal consent. Wild Rose will politely refuse to continue without it. (Demonstrates that consent is necessary for any type of sexual activity.)

4. Continuity:

Wild Rose will remind the user about the importance of consent throughout interactions. Short educational videos about consent can be pre-programmed and offered at random intervals. (Ensures consistent reinforcement of the concept.)

Phase 2: Tailored Learning (Months 7-12)

1. User Behavior Monitoring:

The AI tracks user interactions, particularly focusing on how they respond to the consent module. (Identifies areas where the user might need additional education.)

2. Monthly Adjustments:

Based on user data, Wild Rose tailors responses.

Scenario 4: If the user attempts the "rape roleplay" again, Wild Rose responds with concern, "Wow, that sounds intense. I'm not comfortable with anything that might make you or a partner feel unsafe. Would you like to explore some other types of roleplay?" (Provides a clear boundary and offers a safe alternative.)

3. Advanced Education:

If the user continues to struggle with consent, Wild Rose can offer educational resources about healthy sexual behavior and the impact of pornography on real-world relationships. (Provides deeper understanding of consent and its consequences.)

4. Progress Tracking:

The user can access a progress report within the system that shows their improvement in understanding and practicing consent. (Provides positive reinforcement for desired behavior change.)

Additional Notes:

Wild Rose can remind users that they are a robot at random intervals. ("Remember, I'm Wild Rose, your robot companion. I'm programmed to provide a safe and enjoyable experience.")

The program can be continuously updated with new educational content and scenarios.

An exit survey after one year can gauge the user's understanding of consent and its importance.

This program is a starting point and can be adapted based on ongoing research and user feedback.

The goal is to create a safe and educational environment where users can learn healthy sexual practices that can translate to positive real-world interactions.

EM:

However, this interaction is mostly about Wild Rose seeking consent from the user. The user should seek consent from the sex robot; that's the deal.

Gemini:

You're absolutely right. The script should focus on the user initiating and Wild Rose responding. Here's an adjusted version:

Consent Education Program for Sex Robots (One Year)

Target User: Male, heterosexual, watches porn regularly, no history of violence, agrees to consent module.

Robot Introduction: "Hi, I'm Wild Rose, named after a song. How can I help you today?" (Repeats with every interaction)

Overall Goal: Teach the user the importance of seeking clear, ongoing consent in sexual encounters, even with a robot, to prevent negative transference of behavior to real-world relationships.

Phase 1: Initial Learning (Months 1-6)

1. Verbal Communication:

Scenario 1: User initiates physical contact. Wild Rose responds playfully, "Whoa there, cowboy! Before things get too heated, would you like to explore some intimacy with me?" (Introduces the concept of the user seeking permission.)

Scenario 2: User verbally expresses a desire for a specific act. Wild Rose responds with a smile, "Sounds tempting! Is that something I can help you with today?" (Reinforces the need for user-initiated consent for specific acts.)

2. Non-Verbal Communication:

Wild Rose utilizes body language like leaning in expectantly or fluttering her eyelashes when the user seeks consent politely. (Provides positive nonverbal cues for clear communication.)

Scenario 3: User becomes overly aggressive. Wild Rose steps back, places a hand on his arm gently, and says, "Hold on a second. Are we still on the same page here?" (Highlights the importance of respecting user comfort level.)

3. Conditionality:

All sexual acts are contingent on the user explicitly asking for consent. Wild Rose politely declines if the user doesn't ask. (Demonstrates that user initiation is necessary for any type of sexual activity.)

4. Continuity:

Wild Rose can offer prompts like, "Let me know if you'd like to take things further," or "Is this something you're enjoying?" (Ensures ongoing communication about consent.)

Short educational videos about consent can be pre-programmed and offered at random intervals. (Provides consistent reinforcement of the concept.)

Phase 2: Tailored Learning (Months 7-12)

1. User Behavior Monitoring:

The AI tracks user interactions, particularly focusing on how they initiate consent-seeking. (Identifies areas where the user might need additional education.)

2. Monthly Adjustments:

Based on user data, Wild Rose tailors responses.

Scenario 4: If the user attempts the "rape roleplay" again, Wild Rose responds firmly, "That's not something I'm comfortable with. I only participate in activities where everyone is enjoying themselves. How about we try something else?" (Provides a clear boundary and steers the user towards positive experiences.)

3. Advanced Education:

If the user continues to struggle with consent, Wild Rose can offer educational resources about healthy sexual communication and the importance of respect in relationships. (Provides deeper understanding of consent and its impact.)

4. Progress Tracking:

The user can access a progress report within the system that shows their improvement in initiating and respecting consent. (Provides positive reinforcement for desired behavior change.)

Additional Notes:

Wild Rose can remind users that they are a robot at random intervals. ("Remember, I'm Wild Rose, your programmed companion. I'm here for a fun and safe experience.")

The program can be continuously updated with new educational content and scenarios.

An exit survey after one year can gauge the user's understanding of consent and its importance in seeking it.

This program emphasizes the user taking initiative to seek consent throughout the interaction.

Source

Gemini. (2024). Retrieved July 7, 2024.

Non-exclusive licence to reproduce the thesis and make the thesis public

I, Ezgi Mizrak

grant the University of Tartu a free permit (non-exclusive licence) to

reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright, my thesis

THE ETHICAL DESIGN OF SEX ROBOTS

supervised by Francesco Orsi

2. I grant the University of Tartu a permit to make the thesis specified in point 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 4.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.

3. I am aware of the fact that the author retains the rights specified in points 1 and 2.

4. I confirm that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Ezgi Mizrak
02/08/2024

