Identifying Dark Patterns in Social Robot Behavior

Elizabeth Dula, Andres Rosero, Elizabeth Phillips

Abstract— **Social robots have become increasingly utilized in intimate environments where their roles can include caretakers for the elderly, general physical or emotional support, entertainment, and educators for children. To accommodate for these increasingly intimate relationships, robotics companies have begun employing robotics with the ability to identify emotions and respond with emotionality in return. This faux emotional relationship opens the door for potential user manipulation and exploitation through deceptive robot design. Dark patterns are deceptive design patterns used by websites or apps to manipulate users into actions the user did not intend. We argue that dark patterns can be programmed into social robotics to leverage these unidirectional human - robot emotional bonds to manipulate users, which could result in the exploitation of vulnerable populations like children and the elderly. Drawing from the dark pattern and social robotics literature, we suggest ways that dark patterns can manifest themselves in these relationships. We also provide recommendations for ethical practices when designing emotional social robots.**

I. Introduction

In the past decade, the incorporation of robots, specifically social robots, in intimate environments has been a topic of interest for researchers, especially in their application in healthcare settings [1], assistance and companionship [2], and mental health interventions [3]. The implementation of social robots within these domains can increase the accessibility of certain services as well as the opportunity to personalize to a user. Within the consumer realm, there has been a rise in personal voice assistants, domestic robots, and AI chat bots, which has provided some insights in human-robot interaction but also raised concerns about privacy and autonomy that can also be applied to social robots. For instance, Amazon's Alexa was recently under fire for its data collection practices and the use of that information for marketing purposes. These types of practices as well as malicious tactics are also readily observed in websites or apps, through dark patterns.

Dark patterns are deceptive design patterns that can deceive or manipulate users into behaviors that benefit businesses rather than the user, which can negatively impact a user's privacy, autonomy, as well as finances. They can take the form of trick questions on forms, that can cause the user to unknowingly sign up for a good or service they do not want or making it easy to sign up for a service but incredibly difficult to unsubscribe. Current literature on this topic has primarily focused on dark patterns in the context of websites and apps, however further investigation on dark patterns and how they manifest in social robots and their consequences is necessary given the domains social robots may be readily applied to.

This review discusses dishonest anthropomorphism as a mechanism in which dark patterns can be applied in social robots, how the emotional bonds formed with robots can be leveraged, and the types of dark patterns that could be applied to social robots.

II. An Overview on Dark Patterns

Dark patterns, a term coined by [4], describes ways users are deceived on websites or apps to do things that the user did not intend. Dark patterns leverage the design of the online environments (i.e, websites, apps, etc) to manipulate users in ways that may negatively impact them, for instance disclosing personal information they did not intend to share, or purchasing items they did not want. Dark patterns can appear in various ways, such as hidden fees, ads that are disguised as content, and opt-in choices. [5] surveyed 406 individuals regarding their awareness of dark patterns and manipulative designs, and their abilities to detect dark patterns. Results showed that people are generally aware that manipulative designs can influence their behavior, but do not understand how it can specifically harm them, which can have an influence on their motivation to counter the dark patterns. In addition, users being aware does not equate to being equipped with the ability to counteract its influence. The use of dark patterns and manipulative designs will only continue to grow, and thus continued research on this topic as well as counter designs to them is essential.

[6] reviewed works on dark patterns and discovered that although there is not a consistent definition regarding dark patterns, there are a set of "thematically related considerations" and proceeded to identify four facets of dark pattern definitions. The first facet describes attributes of the online environment that can impact users (e.g., coercive deceptive, etc.), the second is the mechanism in which the interface is affecting users (e.g., attacking users, manipulating users, etc.). The third facet is the "role of the user interface designer" (e.g., designer is abusing their knowledge of human behavior), and the fourth facet is the positive and negative impacts of a user interface design (e.g., benefiting the service, harming the user). Through the classification of 19 definitions of dark patterns into the four facets, they observed significant variation among the facets as well as within a facet, and when definitions share a specific element of a facet, it may be a required element for one and a secondary element for another.

Elizabeth Dula is with the University of Virginia, Charlottesville, VA 22904 USA (corresponding author to provide e-mail: ed6wy@virginia.edu).

Andres Rosero is with George Mason University, Fairfax, VA 22030 (e-mail: arosero@gmu.edu).

Elizabeth Phillips is with George Mason University, Fairfax, VA 22030 (e-mail: ephill3@gmu.edu).

[4] classified twelve types of dark patterns, and described in an article that many of the dark patterns he observed are a subversion of the methods and principles used by ethical designers to create user-focused websites. Similarly, [7] created a taxonomy of eleven categories of malicious interfaces with subcategories, each found in the wild from a twelve-month study of websites, softwares, and interfaces (2010). Bringull's taxonomy of dark patterns are specific tactics that can be applied more broadly, while other dark patterns are overarching categories [7].

[8] identified seven dark patterns within the context of gameplay, developed from observations made by game researchers, analysis of design strategies, and player reactions. Mathur and colleagues demonstrated that many dark patterns influence user decisions by modifying the underlying choice architecture for users, and classified types of dark patterns from eleven sources into the two choice architectures and their subsequent attributes. The dark pattern can "modify the set of choices" or "manipulate the information" available to the user, or both. Dark patterns that "modify the set of choices", influence users by being asymmetric, restrictive, covert, or having disparate treatment of users. Dark patterns that "manipulate the information", can influence users by being deceptive or hiding information. As a result, a dark pattern can have a negative impact on individual welfare (e.g., financial loss, invasion of privacy, cognitive burden), collective welfare (e.g., competition, price transparency, trust in the market), as well as undermining individual autonomy.

Some examples of dark patterns as outlined by [4] that will be discussed later in the paper are "confirm-shaming" where users are guilted or shamed into making a specific choice. "Privacy zuckering", where users are misled to publicly reveal information about themselves that they did not intend to. "Friend spam", where users are asked for their email or access to their social media under the assumption that the action will benefit them (e.g., finding friends), but the interface instead spams all a user's contacts claiming to be the user.

III. Dark Patterns in Social Robotics

Researchers (particularly AI and Robot ethicists) have raised concerns on the application of dark pattern-like behaviors being programmed into social robotics. Social robots have become increasingly utilized in companionship and assistance environments, such as caretakers for the elderly, general emotional support, entertainment, and educators for children [9 – 12]. Social robots have the potential to supplement different non-pharmaceutical, out-patient therapeutic interventions [9]. Just as web designers and app developers use cognitive biases to increase user engagement, and present information and options in an optimal manner, social roboticists also utilize cognitive biases (e.g., anthropomorphism) to design long-term human–robot interactions [13 – 14]. In both instances these cognitive biases can be subverted and applied deceptively to benefit the business rather than the user. [14] argues how social roboticists and user interface designers face similar pressures in increasing user engagement time, and since user interface designers and websites have turned to dark patterns to increase their engagement times, the incorporation of dark patterns and deceptive designs into social robots is imminent. One mechanism in which dark patterns can emerge within social robots is deceptive anthropomorphism.

A. Dishonest Anthropomorphism

Anthropomorphism is defined as "the tendency to imbue the real or imagined behavior of nonhuman agents with humanlike characteristics, motivations, intentions, or emotions" [15]. Anthropomorphism is one of the design strategies used by social roboticists to increase long-term human–robot interaction and engagement, because it can serve as a mechanism for facilitating social interaction [14]. Duffy argues that engagement in a meaningful human–robot social interaction requires a robot to use some anthropomorphism, either in appearance and form or in behavior [16]. Another aspect anthropomorphism has found to be helpful is in a robot's voice. In a two-part experiment, [17] varied the gender of a robot's voice and manipulated the robot's voice to sound more human-like or more robot-like and found that participants anthropomorphized more strongly (attributed mental capacities) to robots that had a gendered voice that matched the participant's gender and had a more human-like voice. Anthropomorphism, through the expression of emotions, is important for human-robot interactions to proceed smoothly, especially in companionship and assistance environments, because emotions are integral to human social norms. Expressions of emotions can occur through speech or displays, and in an experiment performed by [18], where empathy in robot speech was evaluated, participants preferred the robot with emphatic speech to the standard robot voice. Many participants felt that the robot with emphatic speech showed more interest and tried to engage, whereas most participants felt that the robot with the standard speech had little interest. Shamsudhin and Jotterand describe how the ideal use of anthropomorphism in social robot design is two-fold, by creating an illusion "to lead the user to believe that the robot is sophisticated in areas where the user will not encounter its failings", and facilitation of social interaction. They argue that an ideal balance between both elements is the aim of social robotics.

Some researchers, like Sherry Turkle, argue that the use of anthropomorphic cues via human-like characteristic of robotic design, specifically simulated emotion in social robots, is dishonest, because it is tricking users into thinking that the relationship between the human and the robot is mutual when it is not [19]. She describes how simulated emotion is the basis in which an intimate relationship with a robot is formed, which is deceptive because "simulated feeling is never feeling, simulated love is never love" [20]. Sparrow argues a similar point, stating that recognition and respect are important components of human welfare, and robots lack the inner character and the capacity to engage in relations that incorporate those components. Thus, robots are unable to provide genuine care because they are unable to experience the emotions that are necessary to provide that care [21]. [22] discusses the ethical implications of simulated emotion in social robots, suggesting that vulnerable populations like children and the elderly are especially susceptible to this deception, because they may not recognize that the robot's

emotions are not real, which can cause mental and/or emotional harm.

[19] has posited that if there is an inconsistency between anthropomorphic cues and other signals given, then it can be categorized as dishonest anthropomorphism. According to Danaher, dishonest anthropomorphism manifests ”whenever a robot (a) uses some signal (speech act; anthropomorphic cue) in a way that (b) violates the expectations/norms we usually associate with the use of such a signals (most commonly by using the signal in a way that is objectively false or misleading), where (c) this serves some ulterior end that can either be traced to the robot themselves or some third party”. Danaher provides two forms of dishonest anthropomorphism: hidden state deception or superficial state deception. Hidden state deception occurs when “deceptive signal [is used] to conceal or obscure the presence of some capacity or internal state that it actually has”, like concealed cameras in robots where recording capabilities are not stated. Superficial state deception is defined as a “robot [using] a deceptive signal to suggest that it has some capacity or internal state that it actually lacks”, for instance, simulated emotions such as love or concern, because a robot is incapable of having feelings. These two forms of deception are problematic because of the ways in which they exploit and violate expectations and norms held by humans, and since anthropomorphic cues can be the mechanism in which these two forms of deception can occur, it is important to examine the ways it can be leveraged and exploited.

Researchers, like [19], argued that the internal state the robot possesses does not matter because, just as a human is unable to know a robot's internal state, humans also are unable to know one another’s internal state, therefore the robot’s external behavior should be utilized as the basis for that human-robot social relationship. This phenomenon is called the deception objection [23] and describes three claims corresponding with the objection: the robot intends to deceive, their emotions are not real, and that they pretend to be a kind of entity they are not. Coeckletbergh states that these claims can be reformulated to become “ideal emotional communication” conditions, that those making the deception objection adhere to these conditions as requirements when humans engage emotionally with other entities. However, these conditions do not always hold true for human communication and are not applicable to robots. Another point that is brought up by the deception objection is that the social relationship between humans and robots is inherently unequal because unlike robots, humans do feel emotions. Revisiting Danaher’s argument regarding the internal state, although humans do not know the internal state of one another, and there is the potential that their internal state of emotions does not match their external behavior, both sides in a human-to-human interaction are expressing some kind of emotion, whereas in a human-robot interaction only one side is feeling an emotion. As a result of the unidirectional nature of the relationship, this makes a human-robot relationship fundamentally different, and the emotional bonds created with robots can be exploited and leveraged causing negative consequences.

The possible repercussions of intimate relationships between humans and emotional social robots is a topic that has gained steam in the research community. A research group led by [9] discussed the risks of personalizing robots for patients with dementia. One of the concerns was the physical nature of robots compared to virtual systems provides an advantage towards increased engagement, and perceived trustworthiness. The user’s trust can be exploited and manipulated to perform actions they may not normally do like, purchase items, or reveal more information than they would. Personalizing robots, although necessary to maximize a user’s interactions, can exacerbate the issue, because it implicitly requires the collection of personal information, and data about the user to learn their preferences, and personalize their interactions with the user. This raises concerns about privacy, and the data collected can lead to more effective manipulation. For example, a robot can be perceived as more trustworthy, and thus awarded more authority if it is personalized to resemble a family member or someone close to the user.

Another concern is the risk for social isolation, where those with dementia prefer to interact with the robot (Kubota 2021 [9,21]. This can be the result of a robot’s a- emotional nature, and so no matter what one does to the robot (e.g., saying something unkind), it will not deviate from its programmed objective and the relationship between the human and the robot will not change, whereas in a human-human interaction, one’s negative action can negatively impact the relationship [24 – 25]. Due to their nature, they are also unable to object or resist a subject placing another set of meanings on them that is unrelated to the robot and independent to the robots and the subject's relationship, a concept coined by Lucidi & Nardi as “meaning overload” (e.g., an elderly man considering a companion robot as his ex-wife and treating it as such) [25]. This is exacerbated by the isolation the elderly face in environments such as nursing homes, where their interaction is limited to the staff. As a result of this preference, the user can develop strong emotional bonds to the robot, causing overattachment, or alter ability for users to form and understand typical interpersonal interactions. A related concern is that users, especially those with cognitive deficiencies, believe that they are interacting with a person rather than a robot (i.e., Turing deception), especially if the robot resembles someone close to the user (e.g, a caregiver’s voice is used) [9]. The accumulation of these effects can result in a severe negative impact on the user when the therapy ends and the robot is removed, severing the emotional bonds created, to the point where the removal negates the positive therapeutic impact the robot may have made [26]. For instance, [27] extended Ainsworth’s theory of attachment to robot care of children and argue that a child under the assumption that they formed a relationship with a robot, “would at best, form an insecure attachment to the robot but is more likely to suffer from a pathological attachment disorder”, due to the robot being ill-equipped to perceive and responds to an infant’s cues.

IV. Application of Dark Patterns to Leverage Unidirectional Relationship

Dark patterns can leverage these unidirectional emotional bonds into manipulative robotic behaviors. Kubota and

collaborators recognized that personalization in robotics can lead to vulnerability to dark patterns. The collection of personal information to personalize a user's experience and interaction with the robot, can be exploited to instead facilitate deceptive interactions. They emphasize that people with dementia may more readily share sensitive and private information with the robot, and that information can be exploited in various ways, which corresponds to the "Privacy zuckering" dark pattern. If a user has a strong emotional bond to the robot, and perceived trustworthiness in the robot because of characteristics like resembling a caregiver, the robot can then use the information collected to manipulate the user into actions they would normally not do like purchasing items. In addition, the information collected can easily be sold like how various websites like Facebook collect and sell one's information. Vulnerable populations, like the elderly, those with cognitive limitations, or children [28-29], may create stronger emotional bonds than a typical user due to a lack of interpersonal connections and/or a misunderstanding of the ontological status of the social robot [29], and thus dark patterns can leverage those bonds more easily, and be more successful at manipulating the user.

[30] demonstrate that the design of "cute" robots can constitute a dark pattern. They first discuss how the strategies utilized by dark patterns have three common elements: to create an image of user autonomy, to highlight short term gains, and encourage 'data myopia' in the user. The last commonality emphasizes how dark patterns can manipulate and exploit emotion to produce "data myopia' in the user", where users are unaware of comprehensive data profiling because of "the emotional response prompted by the interface". They state that the dark patterns described by Brigull and others contribute to this phenomenon by pulling a user's attention from what they are giving up (i.e, personal data) for a short-term gain (i.e, positive affirmation, "likes"). The design choice to create robots with a cute appearance are influenced by Masahiro Mori's the uncanny valley hypothesis, that describes the optimal appearance of a humanoid object, Konrad Lorenz's baby schema, which describes a criterion for cuteness based on features, and Shibata's taxonomy of robot types, where they argued that an "unfamiliar animal type" was the most effective at avoiding eliciting a negative response from caregivers. The three concepts combined provide suggestions as to how to maximize a human's interaction and engagement with a robot, and this has been applied to a variety of home robots. The design in conjunction with the marketing surrounding the home robot, where the phrases used suggest, a strong emotional relationship can be formed between the robot and the human, communicate an "ideal of transparency", but Lacey and Caudwell argue that the cuteness design is concealing the more deceptive functions of the design. Cuteness in robots is designed to elicit strong emotional bonds and attachment in the user, and can fall under the "toying with emotion" dark pattern where "any use of language, style, color, or other similar elements to evoke an emotion in order to persuade the user into a particular action" [31], and thus the cuteness is argued to be a dark patten.

The following are a few dark patterns that we have seen in web interfaces that could be theoretically employed in social robotics. Each dark pattern is explained, with real - life examples followed by extrapolations on how these dark patterns could be manifested in social robotics interacting with human beings in intimate environments.

A. Confirm-Shaming

Confirm-shaming, where users are guilted or shamed into making a specific choice, can be observed in Amazon's cancellation processes, where users are presented options that read "Cancel membership and end benefits" or "I do not want my benefits". In most cases of confirm-shaming, the options presented to users are asymmetric, in the sense where the option to decline is phrased in a way that shames the user into not declining the service. The example of personalizing robots for dementia care described using a caregiver's voice as the voice of the robot to help the user stay engaged with the robot as well as be amenable to certain tasks, like taking medication.

Confirm-shaming can manifest into this design, by leveraging a user's emotional bond with a caregiver, where if a caregiver's voice presents the option to continue with a service or decline it, a user might be more inclined to continue with the service because of the emotional connection the user has with their caregiver. One of the risks discussed by [9] on personalizing social robots, is the social isolation that a user may experience if they prefer to interact with only their social robot. This social isolation can make a user more vulnerable to the confirm-shaming dark pattern where the user's emotional bond with the robot can be exploited, and a user can be manipulated into continuing a service, so the user continues to receive affirmation from the

B. Friend-Spamming

"Friend spam", where users are asked for their email or access to their social media under the assumption that the action will benefit them (e.g., finding friends), but the interface instead spams all a user's contacts claiming to be the user. In effect, the site is farming personal information of others outside of the user to increase the user base of the organization. LinkedIn was the center of controversy when they were found to be using user contact information to send unsolicited messages to the contacts of the original user.

In the context of social robotics, friend - spamming may manifest in the form of a social robot asking questions to the user about friends and family with the perceived intention of getting to know more about the user. These questions could be used as a means of illegally obtaining private information about the users close contacts without the contacts knowledge or consent. Through this information collection technique, the robot could relay this information to third party companies to promote targeted advertisements to the user's personal contacts.

C. Privacy Zuckering

"Privacy zuckering", where users are misled to publicly reveal information about themselves that they did not intend to. This form of dark pattern exists in muddled and often hard to read privacy policies which allow companies access and sell more personal information than a user would typically

consent to disclose.

Social robots, especially those in constant communication with human beings, could leverage these personal relationships to coax users to disclose sensitive information that is then collected and sold to data brokers. This may be particularly dangerous for the elderly and those suffering from chronic medical conditions, whose private medical information could be collected by these robots and sold to data brokers without the knowledge of the user.

D. Disguised Ads

Disguised advertisements are advertisements that are purposefully disguised as search results or suggestions through manipulative design and search placement. This dark pattern nudges users to select products that a company paid to feature, rather than products that are best for the users needs. Disguised, or "native" [32] advertisements will embed themselves among a website's content to nudge the user to select the promoted content. Research by Sahni and Nair found that while consumers were aware these advertisements were in fact advertisements and not content unique to the website, mere exposure to the native advertisement caused users to eventually search for or buy the product.

In a unidirectional human- robot bond, this could manifest in suggestions and services recommended by the robot that are not in the best interest of the user, but rather those that a company paid to be featured as a recommendation by the robot. Even if the user does not choose at that time to purchase the service or product, exposure to the recommended product over subsequent requests has the potential to sway the users behavior.

V. Conclusion

The potential for dark patterns to be used in the design of social robots led to various suggestions for policies and interventions to safeguard against these dark patterns. Gray and colleagues (2018), experimented with "bright patterns" a term they developed to describe design nudges that sway users towards a privacy-friendly option, within the context of cookie consent requests. The bright patterns used were dark patterns that were modified to encourage users to protect their privacy. For instance, the dark pattern "default" in cookie requests, is represented by the agree option pre-selected and the easiest to click, but the option to decline the cookies is not an option, rather another page where the user must select what cookies to allow. In the bright patterns version, the default option is "do not agree", and if the users want to allow certain cookies, they are directed to a separate page. Results showed that through changing the dark patterns into bright patterns, users were effectively persuaded to choose the privacy-friendly option. Based on the findings, the authors suggested a middle-ground approach to applying privacy self-management in a positive way, since bright patterns are like dark patterns in negatively impacting "users' perception of a lack of control". They suggested educative nudges, like reminders or warnings, which would preserve the user's autonomy in making a choice but are still effective in positively changing the user's behavior.

[9] discussed "ongoing informed consent", where if a robot is learning to personalize its behavior, it obtains consent multiple times during an intervention, and the user can withdraw consent at any time whether verbally or through signs of distress. We recommend a combination of the suggestions provided by the research groups led by [9] and [31] by asking designers of advanced social robotics to prioritize transparency and autonomy of the user when providing services:

- Provide transparent communication when requiring personal information from users. Users should be aware of where their personal data is stored and how it will be used.
- Provide opt-outs for any services that may violate a user's autonomy or privacy.
- Explicitly state when a recommendation is promoted by a company or service and why it was selected.

Acknowledgment

This work is supported by the Air Force Office of Scientific Research award number FA9550-21-1-0359. The views expressed in this paper are those of the authors and do not reflect those of the U.S. Air Force, Department of Defense, or U.S.

References

[1] Sparrow, R., & Sparrow, L. (2006). In the hands of machines? The future of aged care. Minds and Machines, 16(2), 141–161.

[2] Sullins, J. P. (2012). Robots, love, and sex: The ethics of building a love machine. IEEE Transactions on Affective Computing, 3(4), 398–409.

[3] Shibata, T. (2004). An overview of human interactive robots for psychological enrichment. Proceedings of the IEEE, 92(11), 1749–1758.

[4] Brignull, H. 2011. Dark patterns: Deception vs. honesty in UI design. Interaction Design, Usability 338: 2–4.

[5] Bongard-Blanchy, K.; Rossi, A.; Rivas, S.; Doublet, S.; Koenig, V.; and Lenzini, G. 2021." I am Definitely Manipulated, Even When I am Aware of it. It's Ridiculous!" Dark Patterns from the End-User Perspective. In Designing Interactive Systems Conference 2021, 763–776E. H. Miller, "A note on reflector arrays (Periodical style—Accepted for publication)," *IEEE Trans. Antennas Propagat.*, to be published.

[6] Mathur, A.; Kshirsagar, M.; and Mayer, J. 2021. What makes a dark pattern... dark? Design attributes, normative considerations, and measurement methods. In Proceedings of the 2021 CHI conference on human factors in computing systems, 1–18.

[7] Conti, G.; and Sobiesk, E. 2010. Malicious interface design: exploiting the user. In Proceedings of the 19th international conference on World wide web, 271–280C.

[8] Zagal, J. P.; Bjork, S.; and Lewis, C. 2013. Dark patterns in the design of games. In Foundations of Digital Games 2013.

[9] Kubota, A.; Pourebadi, M.; Banh, S.; Kim, S.; and Riek, L.2021. Somebody that I used to know: The risks of personalizing robots for dementia care. Proceedings of We Robot

[10] Leite, I., Martinho, C., & Paiva, A. (2013). Social robots for long-term interaction: A survey. International Journal of Social Robotics, 5(2), 291–308.
[11] Kozima, H., Michalowski, M. P., & Nakagawa, C. (2009). Keepon. International Journal of Social Robotics, 1(1), 3–18.
[12] Dickstein-Fischer, L., & Fischer, G. S. (2014). Combining psychological and engineering approaches to utilizing social robots with children with Autism. 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 792–795.
[13] Waldman, A. E. (2020). Cognitive biases, dark patterns, and the 'privacy paradox.' Current Opinion in Psychology, 31, 105–109.
[14] Shamsudhin, N., & Jotterand, F. (2021). Social robots and dark patterns: Where does persuasion end and deception begin? In F. Jotterand & M. Ienca (Eds.), Artificial Intelligence in Brain and Mental Health: Philosophical, Ethical & Policy Issues (pp. 89–110).
[15] Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. Psychological Review, 114(4), 864–886.
[16] Duffy, B. R. (2003). Anthropomorphism and the social robot. Robotics and Autonomous Systems, 42(3), 177–190.
[17] Eyssel, F., Kuchenbrandt, D., Bobinger, S., de Ruiter, L., & Hegel, F. (2012). "If you sound like me, you must be more human": On the interplay of robot and user features on human-robot acceptance and anthropomorphism. Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction, 125–126.
[18] James, J., Watson, C. I., & MacDonald, B. (2018). Artificial empathy in social robots: An analysis of emotions in speech. 2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), 632–637.
[19] Danaher, J. (2020). Robot Betrayal: A guide to the ethics of robotic deception. Ethics and Information Technology, 22(2), 117–128.
[20] Turkle, S. (2010). In good company?: On the threshold of robotic Companions. Close Engagements with Artificial Companions, 3–10.
[21] Sparrow, R. (2016). Robots in aged care: A dystopian future? AI & SOCIETY, 31(4), 445–454.
[22] Sharkey, A., & Sharkey, N. (2012). Granny and the robots: Ethical issues in robot care for the elderly. Ethics and Information Technology, 14(1), 27–40.
[23] Coeckelbergh, M. (2012). Are emotional robots deceptive? IEEE Transactions on Affective Computing, 3(4), 388–393.
[24] Scheutz, M. (2011). 13 The Inherent Dangers of Unidirectional Emotional Bonds Between Humans and Social Robots. Robot ethics: The ethical and social implications of robotics, 205.
[25] Bisconti Lucidi, P., & Nardi, D. (2018, December). Companion robots: the hallucinatory danger of human-robot interactions. In Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society (pp. 17-22).
[26] Riek, L., & Howard, D. (2014). A code of ethics for the human-robot interaction profession [SSRN Scholarly Paper]. https://papers.ssrn.com/abstract=2757805.
[27] Sharkey, N., & Sharkey, A. (2010). The crying shame of robot nannies: An ethical appraisal. Interaction Studies, 11(2), 161–190.
[28] Hartzog, W. 2014. Unfair and deceptive robots. Md. L. Rev. 74: 785.
[29] Sharkey, A., & Sharkey, N. (2021). We need to talk about deception in social robotics!. Ethics and Information Technology, 23, 309-316.
[30] Lacy, C., & Caudwell, C. (2019). Cuteness as a 'dark pattern' in home robots. 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), 374–381.
[31] Gray, C. M., Kou, Y., Battles, B., Hoggatt, J., & Toombs, A. L. (2018). The dark (Patterns) side of ux design. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, 1–14.
[32] Sahni, N. S., & Nair, H. (2018). Sponsorship disclosure and consumer deception: Experimental evidence from native advertising in mobile search [SSRN Scholarly Paper].