Assessing the Alignment of Social Robots with Trustworthy AI Design Guidelines: A Preliminary Research Study

Motivation & Background

The evolution of robotics has ranged from basic remote-controlled systems to humanoid robots. With the number of AI features & functionalities increasing for every new A/IS system implemented, security risks too have been making their way as a relevant and significant research topic. There are different kinds of consumer robots, and each one of them has a specific purpose & usage depending on the field of application. Therefore, protecting their systems against possible exploitation or misuse is of utmost importance, and these tasks involves dealing with the specifics of the tech functionalities with each category of robot. Social robots are a specific category of consumer robots intended for social purposes, and most of them are social robots as well. Existing literature suggest that most consumer robots come with vulnerabilities that can potentially compromise the user security and privacy. However, to our knowledge, there are no prior studies on exploration and analysis of tech flaws within the Zumi and Cozmo social robots, and that too in context of robot ethics & trustworthy AI. In this project, we perform a unique research study in which we find tech flaws within Zumi & Cozmo and analyze the impact of these flaws on the alignment of their AI with the standardized IEEE principles for ethics & trust in A/IS.

Experimental Results

It is not hard to imagine why social robots, such as Zumi and Cozmo, which are also social robot-based consumer devices, raise security & privacy concerns. Like some other prior works, our findings show vulnerabilities within them, and how the weaknesses within their system functionalities can be explored leading to security issues and compromise of privacy. However, we have also analyzed these flaws within their AI functionalities in an effort to check and verify their alignment with the IEEE global standardized requirements for ethics and trust. Figure 6 shows our preliminary findings on the two social robots in terms of their alignment with the IEEE standards for ethically aligned, trustworthy AI. The left column represents the flaws analyzed in a robot, and the right column indicates the IEEE A/IS principles that are not fully adhered to according to our findings.

Contribution - Future Scope

- This is a novel research study that finds design flaws within social robots, like Zumi & Cozmo, for determining their ethically aligned, trustworthy AI.
- We will continue our research on other AI systems to further study the alignment of their AI with the IEEE Standards.
- We will conduct user engagement studies to understand how users perceive and react to these design flaws.
- We will explore potential improvements and enhancements to the robots to better align with ethical principles.

References

Anki Cozmo, A Fun, Social Toy Robot for Kids. IEEE A/IS Principles
1. Data Privacy
2. Data Security
3. Human Rights
4. Awareness of Misuse
5. Competency

Cozmo’s inconsistent facial recognition performance, i.e.,
cases of malfunction (or failures) in human face recognition

IEEE A/IS Principles
1. Data Privacy
2. Data Security
3. Human Rights
4. Awareness of Misuse
5. Competency

Zumi’s camera, network access & WIFI connection vulnerabilities (lack of authentication)

Social Robot Design Flaws


Methodology: Analysis Of Flaws

We reviewed and tested different system functionalities of the Zumi social robot, and we found out that there were multiple instances where the system lacked proper authentication. (I) Zumi Network Access & WIFI Connection: We discovered that there are ways in which any user can connect to Zumi and can access it using SSH or Web browser through its WIFI (wireless) connection. Every Zumi robot kit comes with a default login username and password for communication through port 22 (SSH), and through the WIFI connection, where the password is set up as the wireless username i.e. WIFI connection name (for instance, “zumi234” in our experiments). The WIFI connection-based username may vary with each robot kit, but for hacking purposes anyone would just pretty much need to browse the available WIFI network connections and search for Zumi. Thus, it would be easy for the user to access to the WIFI wireless network in the absence of a better security protocol or robust authentication scheme for WIFI access. This is a potential design flaw and can compromise the security of the Zumi system. However, even if a Zumi owner changes the default user password, one may not realize that an additional username can be created and used to access the robot. (II) Zumi Camera Feed: Because of the lack of proper authentication, one can access the Zumi’s built-in camera and hack into its live camera feed. If the Zumi robot is recording via its live camera feed as it is watching, then a malicious user can take advantage of the Monitor feature (within the Zumi camera) and can receive the video stream, thereby covertly capturing the recorded video, and placing the Zumi user privacy at stake. This is another instance of a prospective design flaw, where the system is vulnerable to a successful authentication attack. (III) Zumi Injection Attack: After a successful connection to Zumi through SSH, we were able to perform a short OS command driven injection attack to reveal usernames and passwords on its system. For this injection attack, we created a Python 2.0 script asking a user to input a username and then executed the following code, which injected _shot (on our system)’s system’s cat (echo/paswd), as a test instance, for exploiting Zumi’s lack of authentication as a vulnerability. Figure 5 exhibits the process of this Zumi system flaw-based injection attack. (IV) Cozmo’s Flawed Recognition Functionality: We also extensively tested the Cozmo social robot’s face recognition app that detects, registers and recognizes human faces. We found multiple test case scenarios in which Cozmo is unable to detect correctly and recognize the person successfully (i.e. recall identity. Additionally, Cozmo wrongly detected (or classified) inanimate objects, like statues or video game characters, as seen in Figures 1, 2 & 3, as actual human beings (or real persons). We argue that these instances of anomalous behavior (or malfunction) can mislead users and affect them negatively.