Abdikadar Ali, Danielle Thaxton and Ankur Chattopadhyay
alia3@nku.edu | thaxtond1@nku.edu | chattopada1@nku.edu
NKU - Computer Science Department

NORTHERN KENTUCKY UNIVERSITY

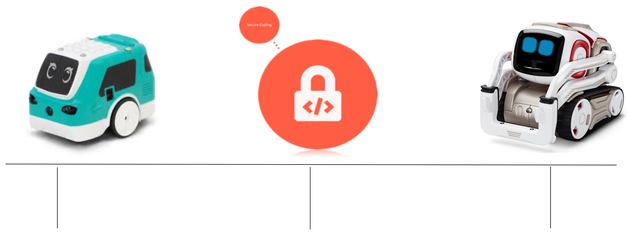
Assessing the Alignment of Social Robots with Trustworthy AI Design Guidelines: A Preliminary Research Study

Abstract

The last couple of years have seen a strong movement supporting the need of having intelligent consumer products align with specific design guidelines for trustworthy artificial intelligence (AI). This global movement has led to multiple institutional recommendations for ethically aligned trustworthy design of the AI driven technologies, like consumer robots and autonomous vehicles. There has been prior research towards finding security and privacy related vulnerabilities within various types of social robots. However, none of these previous works has studied the implications of these vulnerabilities in terms of the robot design aligning with trustworthy AI. In an attempt to address this gap in existing literature, we have performed a unique research study with two social robots - Zümi and Cozmo. In this study, we have explored flaws within the robot's system, and have analyzed these flaws to assess the overall alignment of the robot system design with the IEEE global standards on the design of ethically aligned trustworthy autonomous intelligent systems (IEEE A/IS Standards). Our initial research shows that the vulnerabilities and design weaknesses, which we found in these robots, can lead to hacking, injection attacks, and other malfunctions that might affect the technology users negatively. We test the intelligent functionalities in these robots to find faults and conduct a preliminary examination of how these flaws can potentially result in non-adherence with the IEEE A/IS principles. Through this novel study, we demonstrate our approach towards determining alignment of social robots with benchmarks for trustworthy AI, thereby creating a case for prospective design improvements to address unique risks leading to issues with robot ethics and trust.

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 .



Research 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 Zümi robot kit comes with a default login username and password for communication through port 22 (SSH), and through the WIFI connection route, where the password is set up as the wireless username i.e. WIFI connection name (for instance, "zumi1234" 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 Zümi. Thus, it would be fairly easy way for an outsider to connect to the Zumi 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 Zümi 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 Zümi user privacy at stake. This is another instance of a prospective design flaw plus a security issue, as illustrated in Figure 4. **(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 name, and then executed the following command, which was injected: __import__('os').system('cat /etc/passwd'), 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 responsibly. For example, during our experiments, after registering an individual under a specific name, Cozmo failed to 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.

Experimental Results

It is not hard to imagine why social robots, such as Zümi 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 exploited 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.

Figure 1

Figure 2

Figure 3

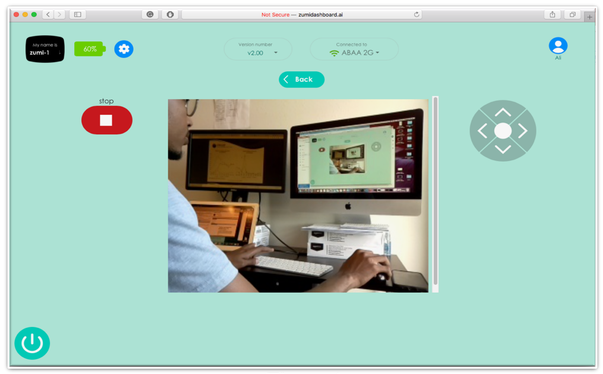


Figure 4

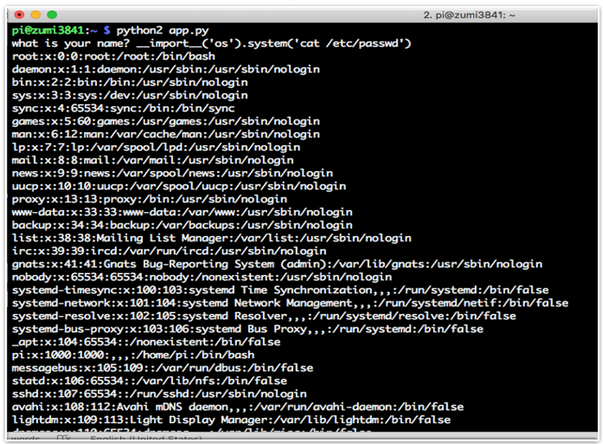


Figure 5

Contributions - Future Scope

- This is a novel research study that finds design flaws within social robots, like Zumi & Cozmo, for determining their alignment with the new IEEE standards for ethics and trust in autonomous & intelligent systems
- Our initial experimental results with the social robots - Zumi & Cozmo, including our analysis of their flaws in the light of the IEEE standards for ethically aligned trustworthy AI, demonstrate a potentially unique approach towards applying the IEEE A/IS Principles for determining adherence with these global design standards
- Our preliminary findings indicate that there are prospective areas of improvement for the tech functionalities within these consumer robots, so that their designs can better align with the IEEE A/IS Standards. Our future work will focus on finding further flaws in other social robots and will continue to analyze these potential design weaknesses in terms of aligning with the principles of robot ethics and trust.

As the volume and sophistication of cyber security attacks increase exponentially, it is necessary to safeguard these social robots and their users from potential exploitation of their flaws, as well as improve their tech feature designs to align with principles and standards for ethically aligned trustworthy AI (in respect of the human users/consumers that these robotic devices are intended for).

Social Robot Design Flaws	IEEE A/IS Principles
Zumi's camera, network access & WIFI connection vulnerabilities (lack of authentication)	1. *Data Agency* *2. Well-Being* *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	*1. Well-Being* *2. Accountability* *3. Effectiveness* *4. Awareness of Misuse* *5. Competency*

References

- "Winner of CES 2019 Best of Innovation Award." Robolink. Accessed Nov. 13, 2020. https://www.robolink.com/zumi/
- "Anki Cozmo, A Fun, social Toy Robot for Kids." Amazon. Accessed Nov. 13, 2020. https://ankicozmorobot.com/
- "IEEE Global Initiative on Ethics of Autonomous &Intelligent Systems." https://standards.ieee.org/industry-connections/ec/autonomous-systems.html
- Kinzler, Matt, et al. "Cybersecurity Vulnerabilities in Two Artificially Intelligent Humanoids on the Market." Workshop on Technology and Consumer Protection (ConPro '19), held in conjunction with the 40th IEEE Symposium on Security and Privacy. 2019 .
- Priyandarshini, Ishaani."Cyber Security Risks in Robotics" ResearchGate, 3 Mar. 2017, www.researchgate.net/publication/319354229_Cyber_security_risks_in_Robotics