Robots, Case-based Reasoning, and ChatGPT: AI-Assisted Ethical Decision-Making

Arkin, R.C., Borenstein, J., Melo Cruz, A.,and Wagner, A.

Determining what is an ethically appropriate decision in day-to-day life can be fraught with significant challenges, especially when vulnerable populations (older adults or children) are involved. Our current NSF-funded project is developing strategies that strive to enable robots to model, generate, display and offer AI-assisted ethical advice. This is challenging in part because formal ethical frameworks such as Consequentialism and Deontology can be at odds and difficult to operationalize from theory to practice. This is further complicated by the need to weigh the importance of folk morality generated from social norms against expert reasoning [1]. What should guide a robot's ethical decision-making process?

We conducted surveys of ethics experts and laypersons to understand what is the right thing to do regarding deception while playing a boardgame with a child, and during pill-sorting training for older adults. We provided a matrix based on demographics and risk, with old-versus-young in one dimension, and high-versus-low task risk for the other.

We then created an architecture using case-based reasoning that generates responses for humanoid robots based on survey data [2]. The goal is to enhance human-robot interaction by maintaining safety on those tasks which may involve significant risk for the user, hence the focus on vulnerable populations in our study. This presentation will discuss survey results and scenarios using the software architecture and robots that we developed.

Finally, we explore ChatGPT, a conversational natural language AI system [3], as a source, in addition to human survey respondents, for identifying what counts as ethical behavior. ChatGPT is an interesting subject of study as it will likely reflect the biases of the data on which it was trained. It is an open question whether ChatGPT may tilt towards being more in line with formal ethical frameworks or folk morality. We explore ethical dimensions of ChatGPT including its potential lack of transparency and explainability, especially in contrast to case-based systems.

[1] Surendran, V., Melo Cruz, A., Wagner, A., Borenstein, J., Arkin, R., and Chen, S., "Informing a Robot Ethics Architecture through Folk and Expert Morality"', *7th International Conference on Robot Ethics and Standards*, Seoul Korea, July 2022.

[2] Chen, S., Arkin, R.C., Borenstein, J. and Wagner, A**.,** "Case-based Robotic Architecture with Multiple Underlying Ethical Frameworks for Human-Robot Interaction", *7th International Conference on Robot Ethics and Standards*, Seoul Korea, July 2022.

[3] https://chat.openai.com/auth/login, accessed 6/19/2023.

This material is based upon work supported by the National Science Foundation under Grant Number 1849068 and 1848974. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.