Conversational Agents for Value Reflection

Michaël N. J. Grauwde[1][0009−0008−4883−2654], Mark A. Neerincx[1][0000−0002−8161−5722], and Olya Kudina[1][0000−0002−9553−5651]

Delft University of Technology: Technische Universiteit Delft m.n.j.grauwde, m.a.neerincx, o.kudina@tudelft.nl
https://www.tudelft.nl/ewi/over-de-faculteit/afdelingen/intelligent-systems

Keywords: Conversational Agents · Human-Computer Interaction · Interactive AI/Human-in-the-Loop Methods and Systems · Human-Centered AI

1 Introduction

In the public safety domain, AI-systems are being developed and deployed for safety measures (e.g. camera algorithmic video surveillance)[10]. The usage of such systems can promote and harm personal values of different stakeholders (such as safety and privacy)[12]. In this paper, we consider how AI-systems may impact values for different stakeholders. We are interested in analysing how conversational agents may aid in reflection on these values to build more value-drive AI-systems. Reflection can be a valuable part of helping us to bring these values to the forefront and can allow us to think about values in a more purposeful manner [13][15][14]. In this study, we look at public safety in AI which allows us to consider the values, tensions between them, and stakeholders' positions. We ask how do people reflect on the implementation of AI in public safety and what values do they consider important in this implementation? We consider that reflection needs (1) a dialogue with both a structure and openness to provide personal content and (2) a conversational "partner" that provides new content for the review. A rule-based agent provides structure and a "prepared/dedicated content", while a LLM-agent provides less structure and "general content" for its dialogue acts. Consequently, we expect a more rich and diverse value reflection with the LLM-agent while we have a paper condition as a baseline condition.

To engage in value reflection, we explore the manner in which conversational agents can be designed for this purpose. Conversational agents can assist people in many different tasks such as behavioural change and compassion [17][11]. They may also have a huge impact in engaging people in reflection in more complex fields such as public safety that involve many different stakeholders that reflect and deliberate on designing and developing new technologies for safety purposes. While they have become increasingly ubiquitous, this is a research gap that exists. In the AI-development pipeline, the values of stakeholders are often insufficiently known, making value-aligned AI-development difficult. We propose that conversational agents provide a unique avenue for non-judgmental reflection. In this work we address: **How can conversational agents be designed to promote value reflection within deliberative processes, specifically in the context of AI-support in public safety scenarios?**

2 Approach – Theoretical Frameworks

We created a text-based conversational agent that attempts to assist people in value reflection in public safety scenarios. In this research, we examine reflection in 3 different stages: 1. Awareness; 2. Understanding; 3. Perspective-taking [1][6][8]. The conversational agent examines this reflection utilising 3 different methodologies: 1. The follow-up methodology [9]; 2. Agonistic Inquiry [2]; 3. Initiating Breakdowns [3][5]. Through engaging people on values, we are able to get a deeper understanding on their decision-making and their thought process. This is a novel approach in the manner that the text-based conversational agent engages people in value-based reflection, in the use of scenarios in the public safety domain to engage value-based reflection [16][4][7] and also in the utilising of different methods with which to engage value-based reflection.

Through using a conversational agent we are able to examine how conversational agents are able to integrate these different methodologies and communicate them with users. We can also gain more insights in the dialogue flow that best leads conversational agents to prompt reflection. A future work will examine integrating state-of-the-art NLU techniques into conversational agents. A continuation of this work will look at how we can build conversational agents in this vein while taking advantage of state-of-the-art Natural Language Understanding techniques building on work by Kocielnik et al. [9]. Particularly intriguing is specifically looking at the manner in which LLMs enhance reflection of the people that interact with them and what are the underlying reason's for this enhanced reflection.

3 Experiment and Study Design

In this research, we consider a within-subjects design between 3 different conditions. Our first condition is 1. A scenario on paper; 2. a text-based conversational agent; and 3. An LLM-based conversational agent. We aim to conduct the experiment with around 60 participants, (20 participants for each condition) to examine the way that the different interventions lead to reflection. In the first condition, participants answer a set of questions after reading a scenario on paper on which they are meant to reflect. The second method places a scenario into a text-based conversational agent that guides the user through a stage-based reflection using different reflection methods. While the last condition is an LLM-based conversational agent that utilises similar methods as the text-based conversational agent while taking advantage of the computational strength of the LLM. Participants are randomly split into one of the 3 groups whose results are then analysed to compare the manner in which the participants have reflected in each condition. The experiment's data will be analysed following our Open Source Framework (OSF) pre-registration.

References

1. Atkins, S., Murphy, K.: Reflection: a review of the literature. Journal of advanced nursing **18**(8), 1188–1192 (1993)
2. Bächtiger, A.: On perfecting the deliberative process: agonistic inquiry as a key deliberative technique. In: APSA 2010 Annual Meeting Paper (2010)
3. Baumer, E.P.: Reflective informatics: conceptual dimensions for designing technologies of reflection. In: Proceedings of the 33rd annual ACM conference on human factors in computing systems. pp. 585–594 (2015)
4. Carrol, J.: Five reasons for scenario-based design. In: Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences. 1999. HICSS-32. Abstracts and CD-ROM of Full Papers. vol. Track3, pp. 11 pp.– (1999). https://doi.org/10.1109/HICSS.1999.772890
5. Dewey, J.: How we think. DigiCat (2022)
6. Fleck, R., Fitzpatrick, G.: Reflecting on reflection: framing a design landscape. In: Proceedings of the 22nd conference of the computer-human interaction special interest group of australia on computer-human interaction. pp. 216–223 (2010)
7. Friedman, B., Hendry, D.G., Borning, A., et al.: A survey of value sensitive design methods. Foundations and Trends® in Human–Computer Interaction **11**(2), 63–125 (2017)
8. Kember, D., Leung, D.Y., Jones, A., Loke, A.Y., McKay, J., Sinclair, K., Tse, H., Webb, C., Yuet Wong, F.K., Wong, M., et al.: Development of a questionnaire to measure the level of reflective thinking. Assessment & evaluation in higher education **25**(4), 381–395 (2000)
9. Kocielnik, R., Xiao, L., Avrahami, D., Hsieh, G.: Reflection companion: a conversational system for engaging users in reflection on physical activity. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies **2**(2), 1–26 (2018)
10. Laursen, L.: A new olympics event: Algorithmic video surveillance (Jan 2024), https://spectrum.ieee.org/paris-olympics-2024
11. Lee, M., Ackermans, S., Van As, N., Chang, H., Lucas, E., IJsselsteijn, W.: Caring for vincent: a chatbot for self-compassion. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. pp. 1–13 (2019)
12. Leslie, D.: Understanding artificial intelligence ethics and safety. arXiv preprint arXiv:1906.05684 (2019)
13. Paxton, J.M., Ungar, L., Greene, J.D.: Reflection and reasoning in moral judgment. Cognitive science **36**(1), 163–177 (2012)
14. Pommeranz, A., Detweiler, C., Wiggers, P., Jonker, C.M.: Self-reflection on personal values to support value-sensitive design. In: Proceedings of HCI 2011 The 25th BCS Conference on Human Computer Interaction. BCS Learning & Development (2011)
15. Prilla, M., Degeling, M., Herrmann, T.: Collaborative reflection at work: supporting informal learning at a healthcare workplace. In: Proceedings of the 2012 ACM International Conference on Supporting Group Work. p. 55–64. GROUP '12, Association for Computing Machinery, New York, NY, USA (Oct 2012). https://doi.org/10.1145/2389176.2389185, https://dl.acm.org/doi/10.1145/2389176.2389185
16. van der Waa, J., van Diggelen, J., Cavalcante Siebert, L., Neerincx, M., Jonker, C.: Allocation of moral decision-making in human-agent teams: a pattern approach. In: Engineering Psychology and Cognitive Ergonomics. Cognition and Design: 17th

International Conference, EPCE 2020, Held as Part of the 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings, Part II 22. pp. 203–220. Springer (2020)
17. Zhang, J., Oh, Y.J., Lange, P., Yu, Z., Fukuoka, Y.: Artificial intelligence chatbot behavior change model for designing artificial intelligence chatbots to promote physical activity and a healthy diet. Journal of medical Internet research **22**(9), e22845 (2020)