An assessment tool for the public opinion of the moral status of Artificial Intelligence

Meghan Hurley[1,2], Gillian Hue[1]

1. Neuroscience & Behavioral Biology (NBB) Emory College of Arts and Sciences
2. Master of Arts in Bioethics (MAB), Center for Ethics, Emory University.

Introduction

Will AI become conscious or sentient?

- Ability to plan actions, integrate and process information in a manner similar to humans
- Increasingly anthropomorphic conceptualization of AI's underlying mechanisms

How will we know? How will we treat them?

- Unclear criteria for defining consciousness and sentience
- Unclear criteria for determining moral status for humans, non-human beings and entities
- Unclear whether moral status confers moral rights or legal rights

Considering that the **public's attitude and acceptance** towards conscious AI will play a large role in deciding how AI are treated and esteemed as members of society or not, it is imperative to understand the public's current opinions of the moral status of AI. This project aims to **develop a robust assessment tool** that can be used to examine factors and themes crucial to the public's opinion of the moral status of AI.

Methodology

Tool Development: to create and evaluate questions that elicited appropriate and authentic responses from the participants**,** scenarios[1,2] were used to combat the limitations of participant self-reporting.[3] Five scenarios were created that assessed themes and subcategories preconceived as being dimensions of moral status

Figure 1. Preliminary concept map of dimensions of moral status

Tool Evaluation: qualitative data collection through four semi-structured interviews

Verbal consent obtained → **Questionnaire administered** verbally (scenarios and demographic / background) → **Semi-structured questions** (thought process and clarity + effectiveness)

Results

Participants: Interviews were conducted with four individuals from a convenience sample. Background and demographic information was collected. Of the 4 participants, 3 were members of the public and 1 was a neuroscience researcher (expert); 2 male 2 female, Ages ranged from 25 to 79.

The interview transcripts were transferred into a separate document, transcripts reread, and themes relevant to AI's personhood and moral status were abstracted from the interview transcripts.

Modification of the concept map of dimensions of moral status relevant to AI:

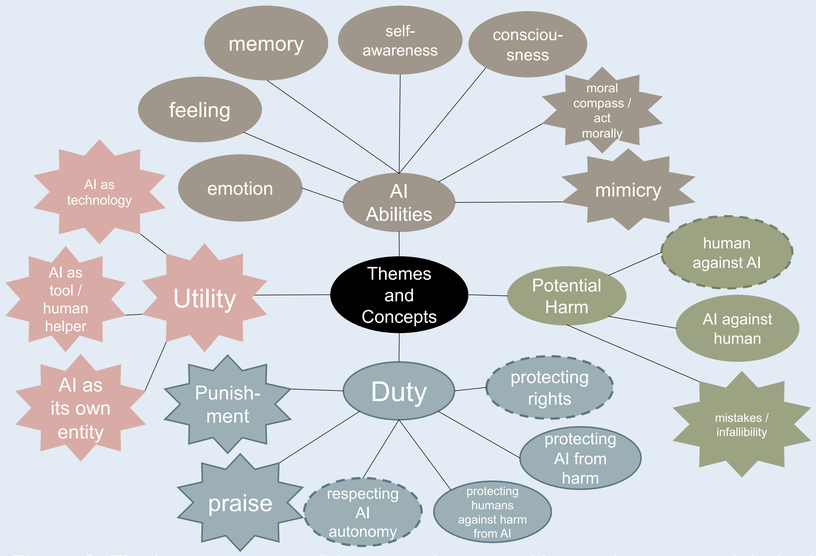

Figure 2. Final concept map of themes relevant to AI, moral status, and their place in society as identified by the study participants.

Four main themes and their respective subcategories comprised the data. Under each main theme, at least one new subcategory was identified through the interview process.

AI Abilities

"I **don't think of AI** as like **having like a moral compass**."

"It comes down to the fact that **it's not human emotion."**

"AI is usually constructed by humans so it's like a human power agency, so **it's like invested with the similar kinds of things that we should value**."

Potential Harm

"I **really didn't care** [what happened to the AI]."

"Even **AI in itself is not infallible… it's almost thinking [like a] human** because same as a human is not infallible either."

"I wouldn't blame [the AI] because it could have been a mistake too… say the AI didn't know it was being manipulated by some outside actor."

Duty

"I thought it **shouldn't be rewarded** for what it should be able to do, or **what it's supposed to do.**"

"The idea of **punishment**, I think it **didn't really fit with my conception of the AI abilities**."

"One **can attribute punishment to AI**… [the answer of] 'don't know' is more attributed to [what type of punishment it should be]."

Utility

[The AI] is just a **piece of technology."**

"I would think of it as its **main purpose is to kind of help prevent human error**."

"I feel like I don't think of the like personal, human-like consciousness aspects [of AI] as being important to me… **I [wouldn't] necessarily benefit** from that aspect of it."

Conclusion

Relevant Themes and Concepts:

- Participant responses provided insight into **further unanswered questions** and **new potential themes** to explore regarding the moral status of AI
- The assessment tool **successfully addressed** four themes— AI abilities, potential harm, duty, and utility— in each of the scenarios
- The **value of this phenomenological assessment** (via interviews) can continue to be explored so that **additional themes and subcategories** can be identified and added to the tool to **ensure its robustness**

Clarity and Effectiveness:

- Feedback **used to modify the questionnaire for future use**
- Two **additional questions added to the demographics and background questions** section and certain questions and scenarios were reworded

Next Steps

Further interviews will be conducted to ensure that 1) a **saturation point of relevant themes** is met, 2) a **wider variety of individuals** with varying understanding of AI is surveyed, and 3) the questionnaire continues to either **directly or indirectly assess** each of these themes. The tool will continue to be modified during this process.

After tool completion, **semi-quantitative and generalizable analyses** will be completed on the public's opinion of AI's moral status.

Author Contribution

MH designed the questionnaire and methodology with guidance and oversight by GH. MH interviewed participants and analyzed the data. All authors approved of this presentation.

References

1. Abu-Odeh, D., Dziobek, D., Torrez Jimenez, N., Barbey, C., & Dubinsky, J. (2015). Active Learning in a Neuroethics Course Positively Impacts Moral Judgment Development in Undergraduates. *The Journal Of Undergraduate Neuroscience Education*, 13(2), 110-119. Retrieved 29 March 2021.
2. Conrad, E. C., Humphries, S., & Chatterjee, A. (2019). Attitudes Toward Cognitive Enhancement: The Role of Metaphor and Context. *AJOB Neuroscience*, 10(1), 35-47. doi:10.1080/21507740.2019.1595771
3. Thoma, S. J., & Dong, Y. (2014). The Defining Issues Test of moral judgment development. *Behavioral Development Bulletin*, 19(3), 55-61. http://dx.doi.org/10.1037/h0100590