Reply
d

Ekaterina Novozhilova¹,*, Kate Mays², Sejin Paik¹ and James Katz¹

¹ College of Communication, Boston University, Boston, MA 02215, USA
² Department of Community Development and Applied Economics, College of Agriculture and Life Sciences, University of Vermont, Burlington, VT 05405, USA
* Correspondence: ekaterin@bu.edu

We would like to thank Dr. Damaševičius for his thoughtful comments on our recent submission [1]. His insights into our approach to measuring public trust and suggestions for further development of the study results are greatly appreciated [2]. We are in agreement that the primary limitation of our paper lies in the shortcomings of the survey methodology, and we acknowledge that incorporating qualitative data and cross-cultural comparisons would significantly enhance our study.

Dr. Damaševičius highlights that the survey approach “may introduce biases related to respondents’ self-assessment of their understanding and experience with AI” [1] (p. 1). We acknowledge that this poses a threat to the validity of our results. However, we implemented several measures to mitigate these potential biases. First, we provided a clear definition of AI at the beginning of our survey: “Artificial Intelligence (AI) refers to computer systems that perform tasks or make decisions that usually require human intelligence. AI can perform these tasks or make these decisions without explicit human instructions” [1] (p. 350). Second, we assessed respondents’ knowledge of AI not only through self-reports but also by administering an AI knowledge test to measure their objective understanding. Respondents were given five randomly selected instances of technologies from a set of 14 and asked to indicate whether those technologies use AI. Each participant’s scores were summed, with a higher score indicating greater AI knowledge. We would further make the point that general survey samples are never truly unbiased. Therefore, future research in this area might want to look at samples of demographic groups that are not prone to participate in online surveys.

The comment suggests introducing a longitudinal research design to demonstrate changes in public trust alongside AI advancements. We agree and recommend conducting a longitudinal study on cohorts (randomized and national samples), which avoids the issue of participant drop-out by not requiring the exact same participants to take the survey each year. While our current work is not a longitudinal study, its results align with previous public opinion surveys we conducted over the recent years with similar samples [1,3]. For instance, our findings that individuals express more apprehension toward AI use in therapy compared to its applications in education and creative domains mirror earlier findings indicating greater comfort with AI teachers, journalists, or surgeons rather than AI counselors [4]. Furthermore, consistent with our previous research, traits such as gender and technological competence continue to predict people’s attitudes towards AI, suggesting these correlations hold regardless of technological advancements [3,4]. However, our approach to measuring trust in AI based on its benevolence and capability revealed an important nuance: individuals with higher education, knowledge, and familiarity with AI were less inclined to perceive it as benevolent in the education domain, yet they still
trusted its capabilities. Given that our survey was conducted shortly after the roll-out of ChatGPT, this shift in attitudes may be attributed to the abundance of negative opinions regarding its implications for education. It would be interesting to investigate whether these attitudes persist, especially considering the recent reduced coverage of ChatGPT.

We agree with Dr. Damaševičius that our study’s results would be enriched by incorporating qualitative research, as noted earlier. We provided participants with the option to elaborate on AI applications in various domains and collected responses from 38 percent of the sample. These responses were reviewed, and some were included as comments that illustrated quantitative findings. We also proposed investigating public attitudes through public assemblies, similar to the approach of van der Veer et al. [5]. In addition to in-depth interviews and focus groups, we recommend using the scenario method, which involves asking participants to write speculative stories about the future impact of AI [6].

Further, we acknowledge the importance of cross-cultural understanding and support Dr. Damaševičius’ suggestion to incorporate views on AI technology from various nations. Recent studies on general attitudes toward AI [7–9] and various AI technologies, such as chatbots [10] and voice assistants [11], reveal significant cultural differences in technology conceptualization, user preferences, and perceptions. However, there are fewer studies conducting cross-cultural comparisons of public trust in AI (e.g., [12,13]), indicating a promising direction for future research.

Finally, we recognize that experimental studies exploring the psychological mechanisms driving trust in AI’s capabilities versus its benevolence would illuminate the causal relationships between these variables. This approach would be particularly beneficial given the proliferation of AI applications for individual use. However, when AI systems are deployed in public domains beyond individual control (such as in public transportation as highlighted in the comment), user preferences and responses may have less influence on policy-making compared to general public attitudes. Our study aimed to demonstrate that while some individuals perceive AI technology as capable and usable, some of its applications in public domains are not necessarily desirable. We argue that compared to the individual users’ studies, our findings of a broad perspective on public trust in AI have greater potential to predict its future adoption and inform the development of appropriate AI policies.

In closing, we once again thank Dr. Damaševičius’ for his thoughtful commentary on our research. Well-informed critical remarks, such as his, serve as a spur to improve the quality of the research enterprise. Hence we are grateful for his commentary and the professionalism that it represents.

**Data Availability Statement:** The datasets generated during and analysed during the current study are available from the corresponding author on reasonable request.

**Conflicts of Interest:** The authors declare no conflicts of interest.

**References**


**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.