

259. TRUST AND MORAL ATTITUDES IN HUMAN–AI INTERACTION

Kuzmina A.A.

*Belarusian State University of Informatics and Radioelectronics
Minsk, Republic of Belarus*

Shevaldysheva E.Z. – Associate Professor

This paper explores how university students perceive and interact with artificial intelligence (AI), focusing on trust, moral attitudes, and willingness to help AI compared to humans. Based on survey data collected from 60 students at the Belarusian State University of Informatics and Radioelectronics (BSUIR), the findings indicate that trust in AI is context-dependent, being higher for technical tasks and lower for personal matters. Moral attitudes toward AI were found to be weak, with the majority of respondents reporting no feelings of guilt when mistreating AI. Furthermore, the willingness to provide assistance was significantly lower for AI than for humans. These results are consistent with recent research published in *Computers in Human Behavior* and have practical implications for the design of AI systems.

Artificial intelligence has become an integral part of daily life for the majority of university students. Whether it is turning to ChatGPT for help with an essay, using a voice assistant to set a reminder, or getting recommendations from a chatbot, AI has become a routine part of student life. These tools make studying easier and faster. They provide students with quick access to information, help them work through complex problems, and even suggest what to learn next based on their needs. However, despite the widespread presence of AI in universities, the question of how students actually perceive these technologies remains understudied. Key questions addressed in this study include the extent to which students trust AI compared to humans, whether they experience guilt when mistreating a chatbot, and how their willingness to help differs between humans and robots. This study investigates these questions through a survey of 60 students from BSUIR. The research examined the levels of trust in AI in various situations, the presence or absence of guilt toward technologies, comparative willingness to help humans and robots, and willingness to correct AI mistakes to improve its performance.

Before moving to the empirical part, an analysis of existing research in this area was conducted. Several studies published in the authoritative journal *Computers in Human Behavior* are of particular interest. One notable study by Volpato, DeBruine, and Stumpf [1] synthesized findings from multiple studies and identified an important pattern. According to their conclusions, people tend to trust AI more when seeking factual information, whereas in matters related to emotions or health, human participation is preferred. This indicates that trust in AI is situational. Research by Li et al. [2] demonstrated that the appearance and behavior of a chatbot significantly influence people's willingness to interact with it. When a chatbot has human-like features – a name, personality, or the ability to express emotions – users are more likely to want to help it. This reaction is explained by the emergence of empathy rather than a conscious moral duty toward an artificial agent. The work of Duan et al. [3] examined the dynamics of trust in mixed teams consisting of humans and AI. It was found that people begin to perceive AI as a full-fledged team member only when it proves to be useful, reliable, and consistent in its results. These studies served as the theoretical basis for conducting the empirical research.

The study employed a survey methodology with data collected from 60 BSUIR students aged from 18 to 22 years. Respondents represented various fields of study: engineering, economics, and information technology. Participation in the survey was voluntary and anonymous, which allowed for obtaining honest answers. The questionnaire was created using Google Forms and distributed through student groups on Telegram and VKontakte. Data collection took place over two weeks in February 2026. The survey questions covered the following aspects: frequency of using AI tools, level of trust in technical and personal contexts, assessment of the fairness of decisions made by AI, moral reactions to hypothetical rude treatment of a chatbot, willingness to help a human and a robot, and willingness to correct mistakes made by AI.

Analysis of the data revealed several significant patterns characterizing the psychological perception of AI among young people. Regarding frequency of use, the vast majority of respondents (55 %) reported daily interaction with AI technologies. Another 33.3 % use AI at least once a week. Only 11.7 % described themselves as rare users. These findings confirm the deep integration of AI into the daily academic and personal students' lives. Trust assessments revealed its pronounced contextual dependence. Thus, 55 % of respondents expressed greater trust in AI recommendations when choosing a technical device (for example, a laptop). At the same time, 58.3 % of respondents would prefer to consult a human doctor when health problems arise. Assessment of AI fairness showed that 50 % of participants agree with the statement that AI

is capable of making impartial decisions. Another 30 % found it difficult to answer, taking a neutral position. 13.3 % expressed disagreement, and only 6.7 % fully supported this statement. The results concerning moral attitudes deserve special attention. The vast majority of respondents (73.3 %) stated a complete absence of guilt in the case of deliberate rudeness toward a chatbot. This category of participants views chatbots exclusively as tools that do not deserve moral consideration. 23.3 % suggested they might feel discomfort, and 3.4 % could not determine their position.

The most revealing results were obtained when comparing willingness to help a human and a robot. The vast majority of participants (86.7 %) expressed unconditional willingness to help a person carrying heavy bags, which reflects ingrained social norms of mutual assistance. However, when presented with a similar scenario involving a robot struggling to move a heavy load, the picture changed dramatically. Only 11.7 % of respondents stated an unconditional willingness to help the robot. The majority of respondents (66.7 %) chose "probably no." 15 % answered "probably yes," and 6.7 % – "definitely no." The data obtained clearly demonstrate that helping behavior remains predominantly human-centered and does not automatically extend to artificial entities. Interestingly, when moving from direct helping behavior to cooperative behavior, a different picture emerges. 81.7 % of respondents expressed willingness to voluntarily correct mistakes made by a chatbot, indicating a pragmatic approach to interacting with AI as a useful tool, despite the absence of moral concern.

The results obtained are consistent with the conclusions of studies published in *Computers in Human Behavior*. The identified contextual dependence of trust confirms the theoretical propositions formulated by Volpato et al. [1] that trust relationships between humans and AI vary depending on situational demands and perceived risks. The widespread absence of guilt when treating chatbots rudely corresponds to the conclusions of Li et al. [2] about the conditions under which people extend moral consideration to artificial agents, namely the need for anthropomorphic features to trigger empathetic responses. The significant discrepancy between willingness to help humans and robots serves as convincing evidence for the theoretical position that prosocial behavior toward AI requires specific emotional triggers and does not occur automatically. Furthermore, the high willingness to correct AI mistakes aligns with the findings of Duan et al. [3] regarding the utilitarian attitude toward AI systems perceived as useful and reliable tools rather than social partners. This consistency across studies strengthens the validity of the present findings.

The findings have practical implications for various groups. For developers and interface designers, the research results indicate that strategically incorporating anthropomorphic elements may increase user engagement and willingness to interact cooperatively with AI systems. At the same time, developers should consider contextual limitations, understanding that in sensitive areas such as healthcare or financial planning, users may consistently prefer human interaction regardless of the level of technological sophistication. For educators, understanding the characteristics of students' perception of AI can facilitate more effective integration of these technologies into the educational process, allowing emphasis on practical applications while acknowledging existing limitations. For companies implementing AI in customer service, the context-dependent nature of trust indicates the advisability of implementing hybrid models in which AI handles routine information requests while maintaining the possibility of seamless access to a human in emotionally sensitive or particularly significant situations.

Several methodological limitations must be considered when interpreting the results. The sample is limited to BSUIR students aged from 18 to 22 years, which does not allow generalizing the findings to older age groups or other cultural contexts where attitudes toward technology may differ significantly. Furthermore, the survey methodology relied on hypothetical scenarios rather than observation of actual behavior during interaction with AI, which could potentially lead to discrepancies between stated intentions and actual actions. Future research could address these limitations by implementing experimental designs involving real interaction with AI systems, systematically varying anthropomorphic characteristics to identify their influence on helping behavior, and including representatives of different age groups and cultural contexts to ensure broader generalizability of findings.

Despite the noted limitations, this study contributes to understanding the psychological aspects of human-AI interaction among university students. Four main patterns emerge with sufficient clarity and consistency: widespread and regular use of AI, confirming the deep integration of technology into student life; pronounced contextual dependence in trust relationships, in which AI is preferred for technical matters and humans for personal matters; an almost complete absence of moral obligations toward artificial agents; and significantly higher willingness to help humans compared to robots, contrasting with pragmatic cooperation with AI as a useful tool. These results are generally consistent with contemporary research in the field of human-AI interaction and contribute to a deeper understanding of the psychological mechanisms through which young people perceive increasingly prevalent artificial intelligence technologies. As AI continues to penetrate various spheres of life, understanding these mechanisms becomes a necessary condition for

creating technologies that will not only be effective but also align with human social and moral norms. This underscores the importance of integrating psychological insights into AI design. Further research should explore how these patterns evolve as AI technologies become increasingly sophisticated and integrated into everyday life. The findings suggest that while students readily adopt AI tools for practical purposes, their willingness to engage with AI on a social or moral level remains limited. Addressing this gap will require not only technological advancements but also a deeper understanding of how users form emotional connections with artificial agents.

References:

1. Volpato, R., DeBruine, L., & Stumpf, S. (2025). Trusting emotional support from generative artificial intelligence: a conceptual review. *Computers in Human Behavior: Artificial Humans*, 5, 100195. – URL: <https://www.sciencedirect.com/journal/computers-in-human-behavior-artificial-humans> (date of access: 15.03.2026).
2. Li, J., Zhu, Z., Zhang, R., & Lee, Y.-C. (2025). Exploring the Effects of Chatbot Anthropomorphism and Human Empathy on Human Prosocial Behavior Toward Chatbots. *arXiv preprint arXiv:2506.20748*. – URL: <https://arxiv.org/abs/2506.20748> (date of access: 15.03.2026).
3. Duan, W., Zhou, S., Scalia, M. J., Freeman, G., Gorman, J., Tolston, M., McNeese, N. J., & Funke, G. (2025). Understanding the processes of trust and distrust contagion in Human–AI Teams: A qualitative approach. *Computers in Human Behavior*, 165, 108560. – URL: <https://www.sciencedirect.com/journal/computers-in-human-behavior> (date of access: 15.03.2026).