Interactions With Smart Machines: Can We Trust AI?

Anica Cvetkovic[a,*]

[a]Faculty of Social Sciences, Tampere University
ORCID (Anica Cvetkovic): https://orcid.org/0000-0002-9555-6241

Abstract. Artificial intelligence is considered to be one of the greatest technological advancements of modern times. Smart machines are expected to transform the ways information is stored, decisions are made, and knowledge is shared. AI is already capable of performing many tasks, and it is used to assist work in different fields such as healthcare, education, entertainment, and business. However, as with any other significant innovation, there is a lot of uncertainty and fear around it. For example, AI is not yet capable of ethical reasoning, and as such it is often perceived as potentially dangerous or unpredictable. Recently, the interest in exploring users' attitudes toward AI has increased, and people's trust in AI arose as one of the key questions. Understanding the concept of human trust in AI is essential for developing safe and reliable machine intelligence. To that end, this study aims to explore human trust in AI while utilizing self-determination theory, self-efficacy, and locus of control as theoretical frameworks. This study relies on the longitudinal survey data. The data consist of information related to people's experiences and perceptions of AI, and also experimental manipulations in which participants are exposed to hypothetical scenarios such as working with AI assistants.

1 Introduction

AI technologies are becoming widely used in various contexts across the globe. AI has not only been integrated into the services and products of businesses and organizations, but also embedded in the gadgets ordinary people rely on daily, such as phones and computers. AI's growing influence on everyday activities and interactions has made the issue of trust a critical concern. Research has shown that trust in AI affects people's behavior, particularly AI acceptance [5] and adoption [1]. Besides this, trust in AI has also been found to affect cognitive and affective processes such as, for instance, opinions and emotions [13]. With trust being a cornerstone of society and AI becoming an increasingly common actor within it, it is important to investigate the role and implications of its growing presence. This has also been recognized by international organizations such as the European Union and United Nations, which are actively working to ensure the safe and trustworthy employment of AI [12, 3].

In the future, AI will undoubtedly become more autonomous, and it will be in charge of a wide range of tasks, from solving complex problems to making independent decisions. Meanwhile, human reliance on smart technologies may continue to grow. Consequently, the issue of trust receives even greater significance. To that end, this thesis aims to investigate human trust in AI through the lens of established theories and concepts from social psychology, specifically self-determination theory, locus of control, and self-efficacy. This sociopsychological perspective offers valuable insights into how individuals can cope and navigate in an increasingly technology-driven society.

* Corresponding Author. Email: anica.cvetkovic@tuni.fi

2 Theoretical Background

There are multiple ways to define trust in the context of human-AI interactions. According to one of the most widely used definitions in the field, trust represents "the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party" [6, p. 712]. This definition originally refers to human-human interactions, but it can also be extended to a human–AI context. In this case, trust represents situations in which an individual believes that AI will complete the task for them without the need to monitor or control it. It is important to highlight that trust depends on many factors. A review of literature has categorized them into factors related to 1) characteristics of a human as a trustor; 2) characteristics of AI as a trustee; 3) the context in which a human and AI interact [4]. The fourth predictor, named interactions, refers to the dynamic aspects of trust in AI, because trust can increase or decrease over time [11].

Locus of control (LOC) is defined as the extent to which one feels to have control over their own behavior, and it can be external or internal [8, 9]. According to Rotter [9], punishments and rewards that we receive from our environment are either dependent on our personal actions and decisions (internal LOC), or they depend on external factors such as luck or fate (external LOC). Self-efficacy refers to one's estimation that they are capable of executing a specific task in a successful way [2]. One of the early definitions in the context of technology implies that self-efficacy can be seen as "the belief in one´s ability to successfully perform a technologically sophisticated new task" [7, p. 467]. Self-determination theory SDT is a concept according to which autonomy, relatedness, and competence represent basic psychological needs that are associated with our well-being and motivation [10]. According to SDT, autonomy refers to one's need to be in control of their own actions and choices; relatedness refers to one's need to establish meaningful connections with others; and competence refers to one's perceptions of their abilities to attain desired results [10].

Because self-efficacy, locus of control, and basic psychological needs are drivers of human behavior and motivation, they also play an important role in shaping people's experiences and well-being in technology-driven environments. Analyzing how these constructs affect trust in AI can be useful for explaining user behavior, in particu-

lar, willingness to interact with AI, accept it, and use it in a way that promotes digital well-being.

3 Research Questions

RQ1: How is locus of control associated with trust in AI? RQ2: How is self-efficacy associated with trust in AI? RQ3: How are basic psychological needs of autonomy, competence, and relatedness associated with trust in AI?

4 Data and Analysis

The study utilizes longitudinal survey data collected in the frame of the UrbanAI (2021-2024), Self & Technology (2022-2025), and WorkSome (2018-2025) projects.

We used the Norstat Finland platform for data collection in all projects except Self & Technology (2022-2025), for which we used both Norstant and Dynata. The data are analyzed quantitatively by employing Stata software. Multivariate analyses are conducted using regression and variance analysis, suitable for longitudinal and comparative analyses.

5 Contributions so far

My doctoral dissertation will consist of four peer-reviewed articles, and I am currently in the process of working on my final article. The contributions made so far are presented below.

In the first study, we investigated whether people have more trust in: 1) AI assistants, 2) human assistants, or 3) AI assistants with human verification, and how control over the assistant affects this trust. In this study, we focused on the workplace context. To investigate this, we conducted an online survey experiment with adult participants from Finland (N = 828). We confirmed our first hypothesis that people trust human assistants more than AI assistants. When it comes to the second hypothesis, we discovered that having control increased trust not only in AI assistants but also in human assistants.

The second study was based on a longitudinal four-wave dataset collected online from adult participants from Finland in 2021 (N = 1226), 2022 (N = 828), 2023 (N = 717), and T4 (N = 653). The study investigated attitudes towards the use of AI in the context of self-determination theory and locus of control. The study looked into attitudes toward AI in twelve domains of AI employment: medicine, caregiving, teaching and education, traffic, urban planning, building and real-estate technology, defense forces, information security, job recruitment, dating services, culture and art, and political decision making. We found significant associations between technological relatedness and positive attitudes towards AI across domains both at within- and between-person levels. Technological autonomy and internal LOC were linked to positive attitudes toward AI only in specific domains, while no link was identified between competence and positive attitudes toward AI.

In our third study, we investigated trust in AI by conducting surveys with participants from 12 countries on 6 continents. This study is grounded in sociopsychological concepts of self-efficacy and self-determination theory. We used linear regression to analyze the data. This study contributes to the discussions on closing the "AI divide" as it includes culturally diverse perspectives. Because this paper is currently under review, detailed descriptions of the findings cannot be provided at this point.

6 Conclusion and directions for remaining work

This thesis is grounded in well-known theories from social psychology and supported by longitudinal data. By analyzing how basic psychological needs, locus of control, and self-efficacy affect trust in AI, it aims to offer novel insights into human behavior and well-being in an AI-driven world. Further, this research responds to the growing call for a more human-centric and cross-cultural approach to studying AI. I am currently working on my final article, which will explore how trust in AI has changed and evolved between 2019 and 2024. As this part of the work is in its early stages, I would greatly appreciate feedback, suggestions, and ideas from fellow researchers participating in this doctoral consortium.

References

[1] T. A. Bach, A. Khan, H. Hallock, G. Beltrão, and S. Sousa. A systematic literature review of user trust in ai-enabled systems: An hci perspective. *International Journal of Human-Computer Interaction*, 40(5): 1251–1266, 2024. doi: 10.1080/10447318.2022.2138826.

[2] A. Bandura. Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84(2):191–215, 1977. doi: 10.1037/0033-295X.84.2.191.

[3] European Commission. Ai act: Regulation 2024/1689. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai, 2024.

[4] A. D. Kaplan, T. T. Kessler, J. C. Brill, and P. A. Hancock. Trust in artificial intelligence: Meta-analytic findings. *Human Factors*, 65(2): 337–359, 2023. doi: 10.1177/00187208211013988. URL https://doi.org/10.1177/00187208211013988.

[5] S. Kelly, S.-A. Kaye, and O. Oviedo-Trespalacios. What factors contribute to the acceptance of artificial intelligence? a systematic review. *Telematics and Informatics*, 77:101925, 2023. doi: https://doi.org/10.1016/j.tele.2022.101925.

[6] R. C. Mayer, J. H. Davis, and F. D. Schoorman. An integrative model of organizational trust. *Academy of Management Review*, 20(3):709–734, 1995. doi: 10.2307/258792.

[7] T. McDonald and M. Siegall. The effects of technological self-efficacy and job focus on job performance, attitudes, and withdrawal behaviors. *The Journal of Psychology*, 126(5):465–475, 1992. doi: 10.1080/00223980.1992.10543380.

[8] J. B. Rotter. *Social learning and clinical psychology*. Prentice-Hall, 1954.

[9] J. B. Rotter. Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs: General and Applied*, 80(1):1–28, 1966. doi: 10.1037/h0092976.

[10] R. M. Ryan and E. L. Deci. *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. The Guilford Press, 2017.

[11] T. Sassmannshausen, P. Burggraf, M. Hassenzahl, and J. Wagner. Human trust in otherware - a systematic literature review bringing all antecedents together. *Ergonomics*, 66(7):976–998, 2023. doi: 10.1080/00140139.2022.2120634.

[12] United Nations. General assembly adopts landmark resolution on artificial intelligence. https://news.un.org/en/story/2024/03/1147831, March 21, 2024.

[13] R. Yang and S. Wibowo. User trust in artificial intelligence: a comprehensive conceptual framework. *Electronic Markets*, 32(4):2053–2077, 2022. doi: 10.1007/s12525-022-00592-6.