

Trust in AI: A Mixed Methods Protocol to Explore Expert Developer’s Trust in Agentic Coding Assistants

Elizabeth Darnell
elizabeth.g.darnell@mytudublin.ie
Technological University Dublin
Dublin, Ireland

Dympna O’Sullivan
Technological University Dublin
Dublin, Ireland

Emma Murphy
Technological University Dublin
Dublin, Ireland

ABSTRACT

As generative artificial intelligence tools are integrated into workplaces, it is critical to understand how individuals trust generative AI tools. We propose a mixed-method protocol to evaluate trust of these tools in the workplace utilising existing trust scales and semi-structured interviews. Trust has been widely studied across disciplines for many years and frameworks of understanding trust have been developed. However, there is a gap in standardised methods to understand and measure trust. The nature of generative AI, in particular the frequency in which it is anthropomorphized, complicates the existing measures and frameworks. This protocol aims to answer questions regarding how accuracy, expertise, and experience impacts trust of these tools, as well as the impact of personal traits. We aim to have this protocol contribute to a greater understanding of trust, especially in the new environment of generative AI, as well as increase the standardisation and replicability of trust research protocols.

ACM Reference Format:

Elizabeth Darnell, Dympna O’Sullivan, and Emma Murphy. 2025. Trust in AI: A Mixed Methods Protocol to Explore Expert Developer’s Trust in Agentic Coding Assistants. In *Proceedings of ACM womENCourage - 12th ACM Celebration of Women in Computing (WomENCourage ’25)*. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/nnnnnnnnnnnnnnnnn>

1 INTRODUCTION

There is an urgent need to understand trust of generative artificial intelligence tools in the workplace as these tools become more ubiquitous across industry. Trust has been widely studied for decades. Yet, it remains difficult to define, measure, and understand. Trust has been found to impact technology adoption which creates a business case to better understand and explore the concept of trust in the era of generative AI [11]. There also exists evidence that ongoing use of a tool can be a valid proxy of trust [3]. Despite these insights, there is no clear understanding of how individuals trust generative AI tools in the workplace and what factors impact that trust. There has been a growing body of work in user-centred AI but the field is nascent and there are few user studies that focus on

trust. This study aims to address this gap via the proposed protocol to evaluate trust that can be applied to diverse settings.

2 BACKGROUND

The gap of understanding of trust exists because there are no widely accepted methods to measure and understand trust [1, 10]. Previous research has introduced the concept of human-human trust and human-machine trust providing frameworks for understanding trust mechanisms in these dyads [4, 8, 10]. It is not currently known which framework of trust generative AI fits best with, especially due to the frequent anthropomorphisation of generative AI [4, 8]. There are additional factors that might impact trust that should be explored, such as accuracy of the tool, trust of the employer and the developers, and individual traits of the end-user. There is not a standard definition of trust that is used across research [1]. The definition of trust we will use is "an attitude that an agent will achieve an individual’s goal in a situation characterized by uncertainty and vulnerability" [6, p.54]. This definition is frequently used and theoretically valid [1, 10]. There has been increased interest and research in trust, particularly in human-computer interaction, but recent research remains primarily exploratory [2, 10]. This work aims to combine both explanatory and exploratory components.

3 RESEARCH QUESTIONS

The research questions that will be explored in this study are:

- (1) How does the accuracy of an agentic coding assistant impact end-user trust?
- (2) How does the end-user’s expertise, demographics, and traits impact trust of an agentic coding assistant?
- (3) Does this proposed protocol effectively measure trust of an agentic coding assistants?

4 METHODOLOGY

4.1 Participants

The proposed study will evaluate the trust of programming experts with an agentic integrated development environment (IDE). Participants will be recruited from academic institutions and industry in Ireland. Recruitment will occur in computationally significant fields such as mathematics, computer science, engineering, and physics. Participants will be proficient with common programming languages like R and Python as well as IDEs. They will be recruited using a snowball sampling method with careful attention to ensure that recruitment has a variety of starting points in the networks to reduce bias [7].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

WomENCourage ’25, September 17–19, 2025, Braşov, Romania

© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM

<https://doi.org/10.1145/nnnnnnnnnnnnnnnnn>

4.2 Rationale for Mixed Methods

In previous trust research, qualitative methods, particularly semi-structured interviews, are heavily relied upon [1]. Scales have also been relied upon, however, they are frequently custom scales and are not validated for application to new populations or contexts [1]. By utilising a mixed method approach for this study, we will be able to utilise the value of the reliability of a standardised scale while including the richness of qualitative data to give further context to the scale, tool, and task. A mixed-method approach to this protocol allows for exploration, development, and validation of new applications of existing methods. Recent literature indicates that there is still a lot to learn about how individuals trust in new contexts, like generative AI, and thus, it is important to retain a significant qualitative component in the protocol [2, 10].

5 PROTOCOL

This study will develop and test a three-phase protocol to measure and understand trust of an agentic coding assistant in the workplace. The protocol will be iterative as it is applied to additional tools and environments beyond this initial setting. The protocol follows a mixed-method approach that utilises an existing trust scale, a coding task with an agentic coding assistant, and a semi-structured interview.

5.1 Phase 1

This phase will be a short survey that utilises the standard trust question, including versions that address trust of generative AI and technology companies. The standard trust question is “generally speaking, would you say that most people can be trusted, or that you can’t be too careful in dealing with people?” [9, p. 72]. Closed-ended questions will be asked about the participant’s use of generative AI tools in the workplace, including which tools they use and what tasks they use and do not use these tools for. At this stage demographics will be collected including age range, gender, nationality, and education level.

5.2 Phase 2

The next phase will be a task with the IDE and the completion of the “Trust in Automation” (TiA) trust scale [5]. Before the task, participants will rate their expertise with a programming language of their choosing via a 5-point scale. The participants will then have 15-30 minutes to explore the IDE. They will be able to define their own prompt or select a prompt from a prompt bank that the research team has developed. Following the task there will be a brief questionnaire about the participant’s perceived accuracy of the generated code and their confidence and trust in that code. Then the participant will complete the 12-point TiA scale [5]. This scale was developed to measure human-machine trust of an automated system which provides the rationale for why it should be evaluated for its ability to effectively measure trust of generative AI tools [5].

5.3 Phase 3

The final phase will be a semi-structured interview that will address a variety of topics including the task itself, the general trust of tool, the participant’s use of generative AI in the workplace, what the role

of generative AI has been in their workplace, and the participant’s general trust.

6 PROPOSED CONTRIBUTIONS

The next steps will be to run this protocol in this context and share results when they are available. After this study has been completed, we will iterate on and refine the protocol based on the results of the study. Given the lack of research in this area, we believe that it is beneficial to develop and share this protocol at this stage allowing for iteration and feedback at all stages. Additional studies will evaluate the trust of retrieval-augmented generated systems that cover workplace policy. We intend for this protocol to be utilised by others in AI research to increase the field’s understanding of trust, increase standardisation and replicability, and improve the methods of measuring trust.

7 ACKNOWLEDGEMENTS

This work was supported by the TU Dublin ARISE (Amplify Research & Innovation Supporting Enterprise) programme, funded under the TU RISE scheme, and co-financed by the Government of Ireland and the European Union through the ERDF Southern, Eastern & Midland Regional Programme 2021–2027 and the Northern & Western Regional Programme 2021–2027.

REFERENCES

- [1] Tita Alissa Bach, Amna Khan, Harry Hallock, Gabriela Beltrão, and Sonia Sousa. 2022. A Systematic Literature Review of User Trust in AI-Enabled Systems: An HCI Perspective. *International Journal of Human-Computer Interaction* 40 (11 2022), 1251–1266. Issue 5. <https://doi.org/10.1080/10447318.2022.2138826>
- [2] Agathe Balayn, Mireia Yurrita, Fanny Rancourt, Fabio Casati, and Ujwal Gadiraju. 2025. Unpacking Trust Dynamics in the LLM Supply Chain: An Empirical Exploration to Foster Trustworthy LLM Production & Use. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA). ACM, 1–20. <https://doi.org/10.1145/3706598.3713787>
- [3] Mathias Bollaert, Olivier Augereau, and Gilles Coppin. 2024. Measuring and Calibrating Trust in Artificial Intelligence. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 14536 LNCS. Springer Science and Business Media Deutschland GmbH, 232–237. https://doi.org/10.1007/978-3-031-61698-3_22
- [4] Francesca Cabiddu, Ludovica Moi, Gerardo Patriotta, and David G. Allen. 2022. Why do users trust algorithms? A review and conceptualization of initial trust and trust over time. *European Management Journal* 40 (10 2022), 685–706. Issue 5. <https://doi.org/10.1016/j.emj.2022.06.001>
- [5] Jiun-Yin Jian, Ann M. Bisantz, and Colin G. Drury. 2000. Foundations for an Empirically Determined Scale of Trust in Automated Systems. *International Journal of Cognitive Ergonomics* 4 (3 2000). Issue 1. https://doi.org/10.1207/s15327566ijce0401_04
- [6] John D. Lee and Katrina A. See. 2004. Trust in Automation: Designing for Appropriate Reliance. *Human Factors* 46 (2004), 50–80. Issue 1.
- [7] Fergus Lyon. 2012. *Access and non-probability sampling in qualitative research on trust*. Edward Elgar Publishing Limited, 85–93.
- [8] P. Madhavan and D. A. Wiegmann. 2007. Similarities and differences between human-human and human-automation trust: an integrative review. *Theoretical Issues in Ergonomics Science* 8 (7 2007), 277–301. Issue 4.
- [9] Eric M Uslander. 2012. *Measuring generalized trust: in defense of the 'standard' question*. Edward Elgar Publishing Limited, 72–82.
- [10] Oleksandra Vereschak, Gilles Bailly, and Baptiste Caramiaux. 2021. How to Evaluate Trust in AI-Assisted Decision Making? A Survey of Empirical Methodologies. *Proceedings of the ACM on Human-Computer Interaction* 5 (10 2021). Issue CSCW2. <https://doi.org/10.1145/3476068>
- [11] Rongbin Yang and Santoso Wibowo. 2022. User trust in artificial intelligence: A comprehensive conceptual framework. *Electronic Markets* 32 (12 2022), 2053–2077. Issue 4. <https://doi.org/10.1007/s12525-022-00592-6>