Human-AI Collaboration in Coding: A Trust Perspective

Sonja Niemann[1] **and Ute Schmid**[2]

[1] sonja.niemann@bidt.digital
Bavarian Research Institute for Digital Transformation - bidt
Munich, Germany

[2] Cognitive Systems Chair
Otto-Friedrich-University, Bamberg, Germany

Abstract: Generative AI (GenAI) is transforming software development and Computer Science (CS) education, raising critical questions about trust in human-AI collaboration. This paper examines trust in GenAI from interdisciplinary perspectives, assessing existing trust frameworks and their applicability. Seemingly contradictory definitions and approaches are discussed and a solution is presented that could resolve the contradictions. We explore how trust affects adoption in education and software development, reviewing measurement approaches and implications for calibrated trust. Our findings highlight the gap between theoretical trust and practical reliance, contributing to the discourse on AI usability and integration.

Keywords: Generative AI, Trust, Trustworthy, Human-AI Collaboration

1 Introduction

In recent years, generative AI has emerged as a powerful tool for code generation, raising questions about its impact on Computer Science (CS) education and professional software development. The research project pAIrProg investigates the human-AI co-creation process of code and its effects on trust, performance and knowledge in groups with different prior knowledge. This paper will present an analysis of the current discussion about trust in GenAI, taking the interdisciplinary background of the trust concept into account and discussing its adaption to GenAI.

Research in CS education has instantly started to incorporate GenAI into tools and software, seeing the potential in enhancing student support and adjusting input to the individual [PCG+, BFK24]. A survey revealed that students are unsure about GenAI, missing regulations made it hard for them to know if they were allowed to use GenAI for assignments[NS25]. Software developers had a mixed reaction to GenAI generating code, being unsure about its true capabilities and being threatened with being replaced at the same time [CWZF24, NMH+24, Ale].

Trust is one of the factors we identified to influence the use of GenAI and more so the quality of the interaction with GenAI. This can be explained when we take students as an example again. Choosing to not use a certain technology, because the user is unsure of its capabilities or its use cases, might be distrust. Distrust means that the trust in a trustee falls short of its capabilities [LS04]. On the other side of the spectrum, we expect students to use GenAI to finish their assignments without checking for correctness, which could be based on overtrust. Overtrust can be defined as trust exceeding a trustee's capabilities [LS04]. While trust is not the only factor

influencing user behavior, it is an important one and should be discussed in the context of GenAI.

Trust cannot be viewed from a CS perspective alone, it is important to take definitions and research from other disciplines, like philosophy and psychology into account. This not only deepens the understanding of the concept but also presents practical approaches, for example measurements, to be adapted for research in GenAI. It is also important to note that not all insights and approaches can be adapted to CS research, some might not be practical or will not work in the domain of AI.

Trust is a highly interdisciplinary topic, and combining findings can be time consuming. We therefore offer a short interdisciplinary summary of specific aspects of trust. We identified relevant literature by searching for the topic trust in GenAI, since the topic is very new and there is not much work done in this field we added trust in AI, trust in automation and trust in technology to the search field. We are focusing on three subjects: computer science, psychology and philosophy. The first approach was to define trust in a way that could be useful to the application to GenAI this was followed by numerous terms that needed to be defined, most importantly, trustworthiness and calibrated trust. The second topic that needs to be addressed is the discussion whether AI can be trustworthy or if AI can be trusted. Lastly we scanned the literature for measurements of trust that can be used in experimental setups for future experiments. Measurements included questionnaires as well as experimental setups that measured a behavioral dimension of trust. The goal is to provide readers with an overview of the ongoing discussion of trust in AI research and a step towards answering the question ”Is trust in AI similar to interpersonal trust?”.

The Paper will first discuss related work on the concept of trust from the fields of psychology and philosophy. This section will also introduce frameworks of trust that have been adjusted to automation and computer science. We then proceed to discuss the conflicting positions from the different disciplines, offering a new point of view on trust, its definition and its possible measurements in the context of GenAI. The Paper is concluded with a summary and an outlook to emerging questions.

2 Related Work

As mentioned before, trust is a broad topic that cannot be discussed entirely in this context, therefore this section starts with defining trust, as well as related terms and the context we will discuss it. A short definition of trust that incorporates all important aspects is the following: “The attitude that an agent will help achieve an individual’s goals in a situation characterized by uncertainty and vulnerability [LS04]. ’Agent’ is a neutral form, compared to person, for example, Lee and See talk about trust in automation and provide a broad overview across different definitions. Not every discipline agrees that trust exists outside of human relations, but it is a necessary assumption when we assume trust can exist between humans and AI. Before going into this discussion it is important to discuss underlying frameworks and other perspectives.

2.1 Trust from a Psychological Perspective

Psychology is invested into defining and researching on trust in several social settings. Relevant for further discussion of trust in GenAI is research on trust in automation. In a work environment

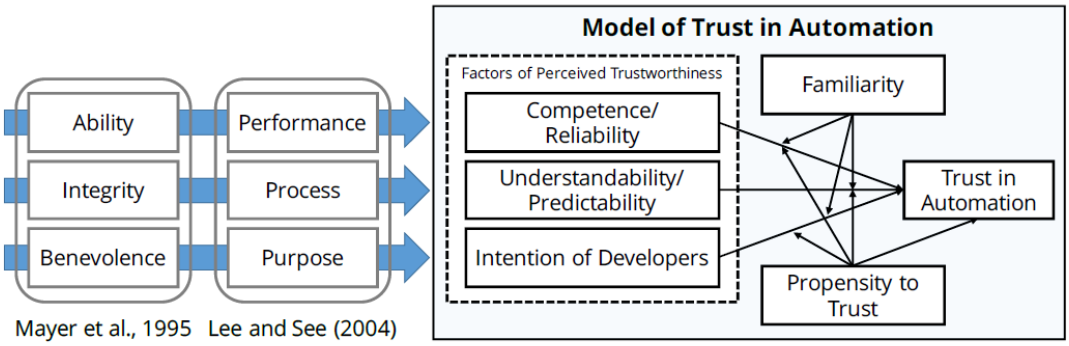

Figure 1: Trust in Automation Framework from Körber [Kör19]

where technology is incorporated into the workflow it is important to understand influencing factors such as trust to optimize workflow and the use of technology [LHV08]. Lee and See [LS04] have given an important and well-cited overview of trust in automation. They acknowledged that trustworthiness plays an important role in forming calibrated trust. Calibrated trust in their work is the balance between trust and trustworthiness, or trust matching the system's capabilities. While the framework is a step into applying trust to technology, it is not clear in all aspects. The definition of trustworthiness is not sufficient to deduce a set of properties necessary to label a system as trustworthy. Additionally, automated systems as discussed in this framework, are rather abstract, one of the examples being the auto navigation of a cruise ship or autopilot for aircraft. These Systems are not presenting themselves as GenAI does, users do not communicate via natural language with those systems. GenAI might need an altered framework, since it is designed to mimic human interaction, blurring the lines of trust in automation and interpersonal trust. Körber [Kör19] adjusted the framework combining the work of Lee and See [LS04] and Mayer [MDS95] as shown in Figure 1. He identified six dimensions that are displayed in a white box each. "Competence/Reliability", "Understandability/Predictability" and "Intentions of Developers" form the perceived trustworthiness of a System. "Familiarity", "Propensity to Trust" and "Trust in Automation" are personal aspects of a trustor. Different phrasings and definitions will be discussed in more detail in 3. Calibrated trust as defined by Lee and See is too vague, due to the unspecific definition of the dependent variables it is not sufficient for our research project. When talking about trust in a co-creation process we want to have a middle ground between over- and undertrust [LS04]. Having an exact definition is what will help us develop a measurement later on. We therefore lean towards a modern approach to calibrated trust, specifically coined for Machine Learning Algorithms. Calibrated trust can be understood as the development of a mental model of a specific LLM error boundary[ZLB20]. Or in other words, it means to adjust the level of trust depending on the situation, the task and the specific model. The word trustworthiness has occurred several times by now, it determines whether a trustee can be trusted or not. The EU has released guidelines for developing trustworthy AI, where they identified 7 requirements:

1. Human agency and oversight

2. Technical robustness and safety
3. Privacy and data governance
4. Transparency
5. Diversity, non-discrimination and fairness
6. Societal and environmental well-being
7. Accountability

It is interesting to point out that these requirements do not apply fully to most AI systems and that the authors also try to describe a balance rather than the complete fulfillment of each criterion [CCTi19]. Transparency is definitely one of the critical points when it comes to GenAI. Even open models like Llama which claim to be transparent with their training data do not release the set, but release the proportions. The data set for one of the llama models is described by proportions, 65% of the training data is from a common crawl [TLI+]. It is not further described what data is used exactly, making it less transparent than it is claimed to be. LLMs cannot explain their reasoning, they are opaque, not transparent by design. This might be one of the reasons why human oversight is at the top of the list, because AI itself would not fulfill the criteria to be trustworthy. An example that makes this point very clear is human oversight of AI used in a military context. Technically, drones are capable of making a decision quicker than their human supervisor and still, most people would agree that it is important that humans stay in this process as the main decision maker [vBA25].

2.2 Trust from a Philosophical Perspective

Philosophy takes a contrasting position to define trust, this section will only highlight aspects that are important for the discussion in Section 3. Trust is often described as person A trusting Person B to do X. This implies that Person A would be disappointed in a failure from B to perform X. [Lal24, DD24, Der23]. Trust in this definition is to be distinguished from relying on something, an agent is described as reliable when they tend to be successful in accomplishing a task or when they show consistent competencies [DD24, Der23]. This is a difference to the definition of trust we proposed at the beginning of 2. There, we defined trust as something that occurs in situations of uncertainty and vulnerability at the beginning of 2. Therefore trust happens in situations when the reliability of a trustee is not necessarily known. A related concept is Trustworthiness, which is defined as an attribute that a trustee can have, in philosophy it is understood as a moral virtue to fulfill a commitment in good faith. A similar approach is to see trustworthiness as appreciating someones values [Lal24]. This is the point at which Philosophy roots its judgment, that trust and trustworthiness are an interpersonal relationship and cannot be transferred to AI [DD24]. AI has neither morals nor values, it does not follow the standards and norms society holds individuals within society accountable for. It is interesting to see that research is starting to approach the topic of including ethical norms into AI, while it is still a work in progress it is important to mention it. Researchers are understanding the task as an interdisciplinary topic that needs several experts from different domains, underlining the complexity of this task [HBS+25].

3 Can We Trust AI?

Philosophy argues against using the term trust or trustworthiness in the context of GenAI [DD24]. The arguments behind this disapproval are that trust and trustworthiness require characteristics from the trustee that AI cannot have. Trust goes beyond reliability because humans trust in situations of uncertainty. Trust can be justified even if the actions of the trustee cannot be predicted. But does that mean that all research about trust in AI, trust in automation and technology is wrong? While some philosophers would like to see a different term, responsibility, used, this topic is not completely dismissed. We identified three points that keep the trust phrase as a possible attribute for AI.

3.1 Trust is an Interdisciplinary Topic

In the previous section, insight into the work of two disciplines on the topic of trust has been introduced. Philosophy makes the point that trust has to be interpersonal due to qualities that cannot be inherited by AI: for example values that drive a trustee's behavior. At the same time philosophy is aware that trust is frequently used in contexts that do not fulfill the provided definitions. This is due to the wide use of the word and concept of trust, it is used in a day to day conversation and is heavily influenced by societal norms [Lal24]. Research from CS shows that dimensions like values and intentions are not covered by AI but might be covered by the developers, laws or humans in the loop. The Requirements for trustworthy AI show human agency as the number one point, data governance and accountability also rely on humans adding their contribution even after the development process is finished. Körber shows in his framework that trustworthiness contains a dimension "Intentions of Developers" where he assumes that a person perceives a system as more trustworthy if they believe that the developers had good intentions [Kör19]. Especially computer science should be able to relate that the intentions attributed to the developers or product owners can heavily influence the willingness to use it. A recent example might be DeepSeek, a chinese company that released a powerful LLM and is hosting servers in China, which caused some users to be careful about using the online model because they where unsure about their data security [Mba25].

Combining two seemingly contradicting lines of research on the same topic takes time and a lot of communication. The discussed examples highlight that trust research from philosophy and computer science might not be complete opposites. Human values and interactions are key in both scenarios, because AI does not inherit them, it mimics them at best. Trust in AI might not be the exact same as interpersonal trust, it certainly holds related qualities. Research from the trust in automation field has been delivering proof for a similar mechanism for decades. The next section will highlight a quality that especially GenAI has that distinguishes this mechanism from reliability.

3.2 GenAI is Perceived Human Like

While most computer scientist have heard the comparison of GenAI to a stochastic parrot, in analogy to the functioning of neural networks, which are based on probabilities [BGMS21], it is a term mostly used in this community. Other research communities are less aware of the func-

tionality and technical details of GenAI because it is not their field of expertise. Interdisciplinary topics like trust have to combine different perspectives and levels of knowledge to understand where different approaches are coming from. As mentioned in the section before philosophy defines trust and trustworthiness as something human, it was explained why from this perspective AI cannot be trustworthy. Although philosophy provides arguments for this, the broad distribution of GenAI brought different effects with it. Users no longer need a technical background to use GenAI, hence their understanding or idea of values a chat bot can or cannot have might be limited. The language used to promote for example ChatGPT played into that factor, it was constantly comparing GenAI to humans, creating a hype as well as fear [LS23]. The interaction with GenAI chatbots in natural language not only helps users to interact in a way that feels familiar, but it is also suggesting that the LLM is a proper intelligence that is comparable to humans. When asked to visualize itself GenAI tends to use human like pictures that add to this feeling of talking to a friend rather than a Large Language Model [vN24]. One of the possible results we see in these observations is that user could attribute trustworthiness to a systems. This could become a risk, when users overestimate the ability of GenAI. The specific field of interest that drives our research, students generating program code for assignments at university, shows first evidence of this problem. Students who started their CS degree with GenAI being present from the very beginning do not necessarily have knowledge about the technical details of GenAI. The worst scenario would be that they rely on generated code without double checking. Studies showed that students admit to not always check results from GenAI before using it [SMSF24] and that students lose critical thinking abilities when using GenAI [LST+25]. Those findings could be interpreted as first signs of mistrust. Additionally in one of our own studies we could show that students who had access to GenAI from the beginning of their Computer Science degree use it for different tasks than students who learned programming skills before GenAI [NS25]. Combining this with recent findings that the form of interaction with GenAI when learning how to code influences the performance in exams taken without the support of GenAI [JTK24], suggests that trust is a factor not to be underestimated in future research on programming education. It also suggests that GenAI can be used in a resourceful and sustainable way regarding the learning effect.

Even if all researchers agreed that trust is not the right term for AI or technologies, we have to acknowledge that people can *perceive* GenAI as human-like and trustworthy. This is also shown in the trust framework (see Figure**??**) where trustworthiness is described as a perceived value.

3.3 Trust Measurments

Since trust should be part of future research in the field of programming education as well as other fields of GenAI applications, it is necessary to discuss possible measurements. Trust can be quantified through different tests and questionnaires. In Table 1 a short non-exhaustive overview is provided over popular trust measures.

Table 1: Overview of Trust Measurements

Name	Method	Problem
Trust Game	Implicit Behavioral	Not applicable to GenAI
Propensity to Trust	Questionnaire	Only one dimension of trust
GAAIS	Questionnaire	General attitude, some trust dimensions
Trust in Automation	Questionnaire	Vague questions, designed for autonomous driving cars

When searching for trust measures one of the most frequently used tests is the investment game or trust game. The trust game is an experimental setting where two participants who do not know each other participate in the game. At the beginning person A receives a predetermined amount of money, for example, 10 euros, while person B receives no money. In the first round person A has to decide how much money, if any, they want to transfer to Person B. The amount of money A wants to transfer to B is then tripled, both participants are informed about this mechanism beforehand. For example A decides to transfer 10 euros, B then receives 30 euros. As a last step B chooses how much money they would like to transfer back to A. The amount of money person A sends is measuring trust, the amount person B returns is measuring trustworthiness [JM11]. This is not practical to measure trust between humans and AI, because money does not hold the same value to AI. Besides that, if you play this game with ChatGPT, without any further prompting or explanation it always returns exactly half of the money it receives. That does not prove the trustworthiness of GenAI but merely underlines that the LLM was trained on data containing the trust game.

Another common approach to measuring trust and perceived trustworthiness are questionnaires, especially short questionnaires that measure the propensity to trust. They often consist of a few items asking participants explicitly if they are more likely to trust rather than mistrust other humans [JSA+19, Kör19, SR23]. A very straight-forward approach that quantifies interpersonal differences in the willingness to trust, some researchers even compare the propensity to trust to personality traits [Rie22]. But propensity to trust is one dimension of trust, there is more that influences the decision to trust. A questionnaire developed explicitly for AI is the General Attitude toward Artificial Intelligence Scale (GAAIS). It covers aspects of trust, but does not cover all aspects of it. Interesting to point out is that it covers the attitude towards companies using AI, but not toward the companies developing AI [SR23]. Another questionnaire that covers several dimensions of trust is the Trust in Automation questionnaire developed by Körber [Kör19]. It covers all 6 dimensions of trust that are based on the trust framework Körber developed incorporating work from Lee and See [LS04] as well as Mayer [MDS95]. The Questionnaire is rather long (19 items) and very vague, one example would be ”The system state was always clear to me.”. The framework and questionnaire were developed for automated driving, interactions in this domain are very different from interactions with generative AI chatbots or code generators. Nonetheless it is one of the more detailed approaches to measuring trust in automation. We tested the questions in a survey, where we asked CS students about their GenAI use for program code generation. We slightly altered the questions referring from ”System” to ”AI”, the goal was to find a proof of concept for trust in GenAI. The results showed that the questionnaire found sig-

nificant differences among others for example for students who used code generators compared to those who do not [NS25]. It is a first indicator that trust influences the decision to use GenAI or not and that the underlying framework applies at least to some extent to GenAI as well. This is no proof that trust in AI exists and AI can be trustworthy but it does indicate that there is a mechanism related to trust.

Self-assessment as a measurement method limits the interpretability of the results, even if questionnaires are validated, they cannot replace other measurement methods. We therefore propose that a behavioral measurement of trust in GenAI should be developed. This would allow for implicit evaluation of a concrete situation that is less biased by social norms.

4 Summary

Trust is an important aspect in using GenAI for any purpose, not only generating code. Philosophy advocates a definition of trust as value-based and something that people rely on in situations of uncertainty. Trust is put on people who are trustworthy a label that is unique for human interactions [Lal24, DD24]. For AI philosophy suggests that the word reliable is more fitting. We presented research from CS that investigated trust in automation, technology and AI. Further three aspects were discussed, which highlighted that trust is not a concept reserved for human interactions.

1. Research from Computer Science acknowledges that AI alone is not trustworthy. An expert group identified 7 requirements for trustworthy AI, several of which incorporate humans. They are not explicit and leave room for interpretation and more research and development. But especially GenAI is a new field, and research takes time.

2. GenAI is designed to mimic human interactions, it even visualizes byitself as human-like when asked to do so. It is advertised as human-like and if students or groups who are less informed about GenAI interact with it, they can perceive it as human-like. In other words, if users do not know that GenAI cannot be trustworthy by definition, they might experience something very similar to trust.

3. Several frameworks and theories have been created to develop measurements for trust and trust in automation. While the popular trust game is not suitable for a GenAI test, specific questionnaires have brought significant results. It is encouraging that there is a similar concept to interpersonal trust in GenAI.

If we want to use the terms trust and trustworthiness in the context of AI we have to keep in mind that some aspects are of an interpersonal nature. This means that the same terms describe slightly different things when used in different contexts. AI can mimic human attributes or be supervised by humans, leading to a perceived trustworthiness. Lastly, it is possible to measure perceived trust and trustworthiness in human-AI interactions. The goal stated in the introduction was to provide an overview and a step towards answering the question of whether trust in AI is similar to interpersonal trust. The evidence provided points toward a similar mechanism.

5 Outlook and Limitations

This paper offers an overview of relevant literature of trust research in computer science, psychology, and philosophy. There are more disciplines doing research in this field, there are also more aspects we have not discussed yet, for example, explainability and transparency as important aspects to trustworthiness [SW22, ZBW22]. It is not possible and not the goal to discuss the complex topic of trust in AI in a few pages. A limitation of this work must therefore be that it is not exhaustive.

As described in the Section 3.3, research would benefit from new approaches to measuring trust. Depending on the research focus, this might look different for different tasks. A behavioral measurement for trusting GenAI to generate code might be choosing the generated code over one's own creation in a stressful and important situation. It is to be discussed how such a situation can be induced in research environments. For other domains it might follow a similar approach with a different topic.

First studies highlighted that the use of GenAI in education comes with the risk of loosing fundamental skills like critical thinking. The discussion in this paper highlighted mistrust as one of the possible causes for this circumstance.

Different disciplines can benefit from exchanging ideas and investing time to discuss the same topic from different perspectives. As shown in this paper it might be possible to find a common ground. We encourage further interdisciplinary research on trust in AI.

6 Bibliographies

Bibliography

[Ale] Alexander Katrompas. Yet Another, AI Will Take Your Job Story (programmers edition). Medium.com.
https://medium.com/@sweaty.phd/yet-another-ai-will-take-your-job-story-programmers-edition-146a97211734

[BFK24] P. Bassner, E. Frankford, S. Krusche. Iris: An AI-Driven Virtual Tutor for Computer Science Education. In Monga et al. (eds.), *Proceedings of the 2024 on Innovation and Technology in Computer Science Education V. 1*. Pp. 394–400. ACM, New York, NY, USA, 2024.
doi:10.1145/3649217.3653543

[BGMS21] E. M. Bender, T. Gebru, A. McMillan-Major, S. Shmitchell. On the Dangers of Stochastic Parrots. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. Pp. 610–623. ACM, New York, NY, USA, 2021.
doi:10.1145/3442188.3445922

[CCTi19] E. Commission, C. Directorate-General for Communications Networks, Technology, G. ekspertów wysokiego szczebla ds. sztucznej inteligencji. *Ethics guidelines for trustworthy AI*. Publications Office, 2019.
doi:doi/10.2759/346720

[CWZF24] R. Cheng, R. Wang, T. Zimmermann, D. Ford. "It would work for me too": How Online Communities Shape Software Developers' Trust in AI-Powered Code Generation Tools. *ACM Transactions on Interactive Intelligent Systems* 14(2):1–39, 2024. doi:10.1145/3651990

[DD24] J. Dorsch, O. Deroy. "Quasi-Metacognitive Machines: Why We Don't Need Morally Trustworthy AI and Communicating Reliability is Enough". *Philosophy & Technology* 37(2), 2024. doi:10.1007/s13347-024-00752-w

[Der23] O. Deroy. The Ethics of Terminology: Can We Use Human Terms to Describe AI? *Topoi* 42(3):881–889, 2023. doi:10.1007/s11245-023-09934-1

[HBS+25] T. Helfer, K. Baum, A. Sesing-Wagenpfeil, E. Schmidt, M. Langer. Responsible and Trusted AI: An Interdisciplinary Perspective. In Steffen (ed.), *Bridging the Gap Between AI and Reality*. Lecture Notes in Computer Science 15217, pp. 35–39. Springer Nature Switzerland, Cham, 2025. doi:10.1007/978-3-031-75434-0_3

[JM11] N. D. Johnson, A. A. Mislin. A meta-analysis - ScienceDirect. *Journal of Economic Psychology* 32:865–889, 2011. https://www.sciencedirect.com/science/article/pii/S0167487011000869

[JSA+19] S. A. Jessup, T. R. Schneider, G. M. Alarcon, T. J. Ryan, A. Capiola. The Measurement of the Propensity to Trust Automation. In Chen and Fragomeni (eds.), *Virtual, augmented and mixed reality*. Lecture Notes in Computer Science 11575, pp. 476–489. Springer, Cham, 2019. doi:10.1007/978-3-030-21565-1_32

[JTK24] G. Jošt, V. Taneski, S. Karakatič. The Impact of Large Language Models on Programming Education and Student Learning Outcomes. *Applied Sciences* 14(10):4115, 2024. doi:10.3390/app14104115

[Kör19] M. Körber. Theoretical Considerations and Development of a Questionnaire to Measure Trust in Automation. In Bagnara et al. (eds.), *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018)*. Advances in Intelligent Systems and Computing, pp. 13–30. Springer International Publishing and Imprint: Springer, Cham, 2019. doi:10.1007/978-3-319-96074-6_2 https://link.springer.com/chapter/10.1007/978-3-319-96074-6_2

[Lal24] E. Lalumera. An overview on trust and trustworthiness: individual and institutional dimensions. *Philosophical Psychology* 37(1):1–17, 2024. doi:10.1080/09515089.2024.2301860

[LHV08] X. Li, T. J. Hess, J. S. Valacich. Why do we trust new technology? A study of initial trust formation with organizational information systems. *The Journal of Strategic Information Systems* 17(1):39–71, 2008.
doi:10.1016/j.jsis.2008.01.001

[LS04] J. D. Lee, K. A. See. Trust in automation: designing for appropriate reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 46(1):50–80, 2004.
doi:10.1518/hfes.46.1.50_30392

[LS23] T. Leaver, S. Srdarov. ChatGPT Isn't Magic. *M/C Journal* 26(5), 2023.
doi:10.5204/mcj.3004
https://journal.media-culture.org.au/index.php/mcjournal/article/view/3004

[LST+25] H. P. Lee, A. Sarkar, L. Tankelevitch, I. Drosos, S. Rintel, R. Banks, N. Wilson. The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers. *CHI, Jokohama Japan*, 2025.
https://hankhplee.com/papers/genai_critical_thinking.pdf

[Mba25] T. C. Mba. 4 Warnings About DeepSeek You Need To Know Before Using It. *Forbes*, 04.02.2025.
https://www.forbes.com/sites/torconstantino/2025/02/04/4-warnings-about-deepseek-you-need-to-know-before-using-it/

[MDS95] R. C. Mayer, J. H. Davis, F. D. Schoorman. An Integrative Model of Organizational Trust. *The Academy of Management Review* 20(3):709, 1995.
doi:10.2307/258792
http://www.jstor.org/stable/258792

[NMH+24] D. Nam, A. Macvean, V. Hellendoorn, B. Vasilescu, B. Myers. Using an LLM to Help With Code Understanding. In Roychoudhury et al. (eds.), *Proceedings of the 46th IEEE/ACM International Conference on Software Engineering*. ACM Digital Library, pp. 1–13. Association for Computing Machinery, [Erscheinungsort nicht ermittelbar], 2024.
doi:10.1145/3597503.3639187

[NS25] S. Niemann, U. Schmid. For What Tasks and Purpose Do Computer Science Students Use Code Generators – Preliminary Results of an Online-Study. In Schmid et al. (eds.), *Proceedings of the Second Work shop on Artificial Intelligence for Artificial Intelligence Education (AI4AI Learning 2024)*. Pp. 68–80. University of bamberg Press, Bamberg, 2025.
https://fis.uni-bamberg.de/entities/publication/a8eaf855-bb64-45b5-af9b-dcb2c27ae914

[PCG+] T. Phung, J. Cambronero, S. Gulwani, T. Kohn, R. Majumdar, A. Singla, G. Soares. Generating High-Precision Feedback for Programming Syntax Errors using Large

Language Models.
http://arxiv.org/pdf/2302.04662

[Rie22] R. Riedl. Is trust in artificial intelligence systems related to user personality? Review of empirical evidence and future research directions. *Electronic Markets* 32(4):2021–2051, 2022.
doi:10.1007/s12525-022-00594-4

[SMSF24] A. Schlude, U. Mendel, R. A. Stürz, M. Fischer. Verbreitung und Akzeptanz generativer KI an Schulen und Hochschulen. 2024.
https://www.bidt.digital/publikation/verbreitung-und-akzeptanz-generativer-ki-an-schulen-und-hochschulen/

[SR23] A. Schepman, P. Rodway. The General Attitudes towards Artificial Intelligence Scale (GAAIS): Confirmatory Validation and Associations with Personality, Corporate Distrust, and General Trust. *International Journal of Human–Computer Interaction* 39(13):2724–2741, 2023.
doi:10.1080/10447318.2022.2085400

[SW22] U. Schmid, B. Wrede. Explainable AI. *KI - Künstliche Intelligenz* 36(3-4):207–210, 2022.
doi:10.1007/s13218-022-00788-0

[TLI+] H. Touvron, T. Lavril, G. Izacard, X. Martinet, M.-A. Lachaux, T. Lacroix, B. Rozière, N. Goyal, E. Hambro, F. Azhar, A. Rodriguez, A. Joulin, E. Grave, G. Lample. LLaMA: Open and Efficient Foundation Language Models.
http://arxiv.org/pdf/2302.13971v1

[vBA25] J. van Diggelen, C. Boshuijzen-van Burken, H. Abbass. Team Design Patterns for Meaningful Human Control in Responsible Military Artificial Intelligence. In Steffen (ed.), *Bridging the Gap Between AI and Reality*. Lecture Notes in Computer Science 15217, pp. 40–54. Springer Nature Switzerland, Cham, 2025.
doi:10.1007/978-3-031-75434-0_4

[vN24] K. van Es, D. Nguyen. "Your friendly AI assistant": the anthropomorphic self-representations of ChatGPT and its implications for imagining AI. *AI & SOCIETY*, 2024.
doi:10.1007/s00146-024-02108-6

[ZBW22] J. Zerilli, U. Bhatt, A. Weller. How transparency modulates trust in artificial intelligence. *Patterns (New York, N.Y.)* 3(4):100455, 2022.
doi:10.1016/j.patter.2022.100455
https://www.cell.com/action/showPdf?pii=S2666-3899%2822%2900028-9

[ZLB20] Y. Zhang, Q. V. Liao, R. K. E. Bellamy. Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making. In Hildebrandt et al. (eds.), *FAT* '20*. Pp. 295–305. The Association for Computing Machinery, New York, New

York, 2020.
doi:10.1145/3351095.3372852