Trust the AI, Doubt Yourself: The Effect of Urgency on Self-Confidence in Human-AI Interaction

Baran Shajari
McMaster University
Hamilton, Canada
shajarib@mcmaster.ca

Xiaoran Liu
McMaster University
Hamilton, Canada
liu2706@mcmaster.ca

Kyanna Dagenais
McMaster University
Hamilton, Canada
dagenaik@mcmaster.ca

Istvan David
McMaster University
Hamilton, Canada
istvan.david@mcmaster.ca

Abstract

Studies show that interactions with an AI system fosters trust in human users towards AI. An often overlooked element of such interaction dynamics is the (sense of) urgency when the human user is prompted by an AI agent, e.g., for advice or guidance. In this paper, we show that although the presence of urgency in human-AI interactions does not affect the trust in AI, it may be detrimental to the human user's self-confidence and self-efficacy. In the long run, the loss of confidence may lead to performance loss, suboptimal decisions, human errors, and ultimately, unsustainable AI systems. Our evidence comes from an experiment with 30 human participants. Our results indicate that users may feel more confident in their work when they are eased into the human-AI setup rather than exposed to it without preparation. We elaborate on the implications of this finding for software engineers and decision-makers.

CCS Concepts

• **Computing methodologies** → **Artificial intelligence**; • **General and reference** → **Empirical studies**; • **Human-centered computing** → *Human-AI interaction.*

Keywords

AI, human-AI interaction, empirical study, self-efficacy, urgency

ACM Reference Format:
Baran Shajari, Xiaoran Liu, Kyanna Dagenais, and Istvan David. 2026. Trust the AI, Doubt Yourself: The Effect of Urgency on Self-Confidence in Human-AI Interaction. In *34th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (FSE Companion '26), July 05–09, 2026, Montreal, QC, Canada.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3803437.3806712

1 Introduction

Recent trends in AI systems have shifted their *modus operandi* from simple stimulus-response systems to more autonomous ones [38]. Mixed-initiative [49] and agentic AI [40] systems, specifically, are autonomous AI systems equipped with the ability to proactively prompt the human, e.g., for guidance, advice, or approval [40]. While evidently of high utility [12], these classes of AI systems introduce novel challenges in human-AI interaction, one of which, as our study shows, is the reduced self-confidence of humans.

Studies that investigate the relationship of human and AI collaborators often focus on trust [30] and its elements, such as the predictability [41] and explainability of the AI agent [2]. Evidently, such properties tend to improve as a result of extended interactions between humans and AI—such as human-AI collaboration [62], guidance [63], cooperation [65] and joint work [17]. As we show, humans' confidence does not necessarily coincide with increased trust. Moreover, **it is possible that trust towards AI increases while the user's confidence (in their own work and role) deteriorates** due to unprepared urgency. We hypothesize that this cognitive asymmetry may lead to the elevated anxiety in users that has been reported in numerous studies [54][26][18].

To guide our investigation, we formulate two research questions.

RQ1. Can we corroborate that interaction with the AI agent improves user trust in our experiment?
First, we aim to establish consistency with the state of the art by demonstrating that in our setup, trust in the AI agent improves in response to human-AI interactions.

RQ2. How does perceived urgency in human-AI interaction affect human users' relationship to the AI agent?
We aim to understand whether and which human attitudes change when subjecting human users to time-pressured interactions with an AI agent.

To answer these research questions, we conducted an experiment with 30 human participants, in which we tasked them with solving an interactive assignment in collaboration with an AI agent.[1] By assessing the participants' trust attitudes before and after the experimental task, we corroborate that trust improves in mixed-initiative AI settings (RQ1)—a positive effect; but we observe that humans react to time pressure with lowered confidence in one's own work (RQ2)—a negative effect.

Our results suggest that AI systems with the ability to prompt humans may have a negative impact on the long-term sustainability of human-AI joint work without careful and gradual introduction of AI tools, and without proper upskilling, education, and training. These implications are important for vendors of AI-based solutions

FSE Companion '26, Montreal, QC, Canada

ACM ISBN 979-8-4007-2636-1/2026/07
https://doi.org/10.1145/3803437.3806712

[1] Replication package: https://doi.org/10.5281/zenodo.19362930.

and organizational leaders alike, as they draw the attention to the need for a coordinated effort when deploying such systems. As well, software engineers as they are responsible for developing urgency cues into system behavior [1].

2 Background and Related Work

2.1 Trust in AI

One of the common definitions of trust is "*the willingness of a party to be vulnerable to the actions of another party (...) irrespective of the ability to monitor or control that other party*" [36]. This is an apt definition in the context of AI systems where black-box systems are typical, and users typically do not possess the ability to assess, monitor, or enforce desirable properties of the system [61]. Consistent with this definition, Zhou et al. [66] conceptualize trust in AI as the expectation that an AI agent will help achieve the user's goal. Hoff and Bashir [22] define trust as the tendency to take risk while believing there is a high chance of a positive outcome.

Elements of trust in AI. To understand the properties and mechanisms through which trust in AI is formed, various decompositions of trust (in AI) have been proposed. In Tab. 1, we aligned four taxonomies of the elements of trust in AI. Clearly, different taxonomies overlap only partially, emphasizing different aspects of trust. Yang and Wibowo [64] and Afroogh et al. [2] conduct systematic literature reviews on user trust in AI sampling from various domains. Nordheim et al. [41] and Seitz et al. [52] focus on trust in chatbots, a particularly interactive AI, conceptually close to collaboration, our topic of interest. We rely on the subset of elements in these taxonomies that are applicable in our experiments. The definitions of those elements are given below. Based on these trust elements, in Sec. 3.2, we assemble a questionnaire to elicit the trust attitude of the participants in our experiments towards AI.

Ease of use The simplicity of technology as perceived by users. This concept is labeled as "usability" in [52].

Dependability A system's consistency over time and across different situations [64]. This is in line with the definition of Avizienis et al. [9]: the ability of a system to avoid constant failure.

Predictability The consistency of the trustee's performance or behaviour in an extended period of time [64].

Alignment The degree of agreement between user- and AI intentions [64]. This is referred to as "benevolence" by Yang and Wibowo [64] and Seitz et al. [52], but we find "alignment" more consistent with recent results in human-AI topics [24].

Transparency The degree to which users understand how AI functions and whether its decision-making process is clear [52].

Table 1: Elements of trust in AI

	Yang and Wibowo [64]	Afroogh et al. [2]	Nordheim et al. [41]	Seitz et al. [52]
Used in our study	Predictability Dependability Alignment*		Predictability	Predictability
		Transparency	Ease of use	Alignment* Ease of use* Transparency
Not used in our study	Ability			Ability
	Faith		Expertise	
		Reliability	Reputation	
		Accuracy		
	Integrity		Risk	Risk Integrity
		Explainability		
				Data Privacy

Asterisk denotes concepts we refer to with an alternative name

2.2 Human-AI Interaction

Mixed-initiative interaction. Mixed-initiative interaction characterizes human-AI interaction as a cooperative process in which both the human and agent contribute at appropriate times to facilitate effective teamwork [4]. Initiative is dynamic rather than fixed, meaning either the human or AI system may act as a leader, supporter, or independently based on the changing demands of the task. Throughout the interaction, roles are negotiated dynamically through a dialogue that determines when the AI system listens versus when it communicates. Mixed-initiative AI systems mark a stark departure from passive AI components that wait for humans to initiate interaction, and instead, they proactively prompt humans.

Agentic AI. Agentic AI is a prime example of AI systems that can operate in a mixed-initiative mode. Agentic AI characterizes autonomous systems that aim to imitate human behaviour to complete tasks with minimal human supervision [40]. The semi-autonomous capabilities of these systems, in addition to reasoning, planning, context-aware interaction, and adaptability, allow these systems to perform complex, multitask problems. Rather than following predefined rules, agentic systems dynamically adapt to changing environments by contextualizing data in different formats, like audio, text, and images. These characteristics position agentic AI as tools that extend beyond automation and toward collaborative task execution, where agentic AI can act as semi-independent assistants to human workflows. Due to the broad scope of these systems, taxonomies categorize these systems based on their level of autonomy, their ability to learn, and their capabilities. These include reactive and proactive agents, limited memory agents, model-based agent, goal driven agents, theory of mind agents, and self-aware agents.

2.3 Related work

Closest to our work are studies that examine confidence and trust dynamics in collaborative human-AI settings. Li et al. [32] investigate the relationship between human self-confidence and AI confidence, finding that human self-confidence shifts to align with AI confidence during collaboration, and that this shift persists even after the interaction ends. An important difference compared to our work is that there is no measure of human *confidence in AI* and that human self-confidence is related to specific problems of a task (e.g., how confidence are you in the label you assigned to this input?). In contrast, our work focuses on users' self-confidence independently of the task at hand (e.g., users are asked about their confidence after playing a game of Pac-Man, rather than being asked to assign their confidence to individual game choices they make).

Park et al. [42] examine how writing with assistance from large language models (LLMs) affects users' self-efficacy and their trust in AI, and how these traits evolve over the course of the interaction. The authors find that LLM-assistance decreased self-efficacy but increased trust in AI. Sullivan and Weger [57] study how varying levels of AI transparency influence users' trust and perceived reliability in AI systems. Their results indicate that higher levels of AI

transparency are associated with higher reported trust and confidence in the system, increased perceived reliability, and improved ease of understanding. Choung et al. [16] explore whether trust influences the acceptance of AI technologies, finding that trust indirectly shapes users' willingness to adopt the technology by increasing perceived usefulness and positive attitudes toward AI.

Ma et al. [34] explore how calibrating human self-confidence influences their confidence calibration, the alignment between an individual's reported confidence and their actual correctness. Their findings suggest that improving self-confidence calibration improves human-AI collaboration and leads to a more rational reliance on AI. These works establish that trust, confidence, and transparency play a key role in improving human-AI interactions. However, they largely examine these attributes without accounting for external pressures that may affect human-AI collaboration, such as time pressure (urgency) in our experiment.

Other works closest to ours are those that examine how urgency affects collaboration in settings between humans and automation. Tatasciore and Loft [58] examine how task time pressure and transparency affect the use of automated decision aids, finding that high time pressure decreases accuracy and increases perceived workload. While they note that increasing transparency improves trust and accuracy, it does not alleviate the negative effects of time pressure. Similarly, Rieger and Manzey [45] investigate how time pressure influences human use of automated decision support systems (DSSs). The authors find that time pressure tends to reduce decision accuracy and that joint human-DSS performance is worse than automation-alone performance. Overall, prior research in human-AI collaboration tends to focus on confidence and trust in low-pressure settings or examines how urgency primarily affects accuracy. In contrast, our work directly investigates how task urgency influences human self-confidence and trust in AI systems, bridging the gap in the previously mentioned works.

3 Study Design

To assess how human-AI interaction impacts trust in AI, we conducted an experiment with human participants, in which we compared participants' trust levels in AI before and after a human-AI interactive task. We recruited 30 participants from various academic levels of computing and software programs, including senior undergraduates and graduate students (Master's, PhD).

We conducted the experiments in person, in a controlled environment at McMaster University, Canada, between October 1 and November 1, 2025. Prior to our experiments, both the environment and the experimental tool were thoroughly tested to ensure reliability and consistency. During each experiment, one participant and the lead researcher were present. We ensured that participants possess the required command in English to avoid language barriers.

3.1 Experimental setup

3.1.1 Experimental tool and task. In our experiments, the participant plays a game of Pac-Man, a popular and well-known arcade game [47]. The participant plays the game in two modes: first, without any AI agent involved; and subsequently, the control is passed to a reasonably trained AI agent that will ask for the participant's advice occasionally. The final score is the combined score of the human-only and human-AI interactive modes. To build emotional investment, the highest scoring player wins a book—a token of modest value, but sufficient to render the participant's stakes in the experiment high enough to test their trust in the AI agent under meaningful conditions.

We developed the Pac-Man game from the Pacman environment of Farama Foundation's Gymnasium framework[2], following established architectural principles of reinforcement learning environments [33]. We trained the AI agent via reinforcement learning using the proximal policy optimization algorithm [50] with the following hyperparameters: learning rate $\alpha = 2.5e{-}4$, discount factor $\gamma = 0.99$, episodes=100, rollout length $n\_steps = 128$. This results in a reasonably trained AI agent that can play Pac-Man.

3.1.2 Experiment overview. As shown in Fig. 1, each experiment consists of an introduction, and two experimental phases.

Introduction The lead researcher *introduces the study* (5 minutes), and answers any questions from the participant. To build investment in the game, the participant is informed that the highest scoring participant wins a prize.

Phase 1: Human-only mode The participant works *without* the AI agent. After a brief *Introduction of Phase 1*, the participant proceeds with a *free game play* (5 minutes), trying to maximize their score. After five minutes of game play, the lead researcher ends the session and records the participant's score. To conclude Phase 1, and before working with the AI agent, the participant fills in a questionnaire about their general attitude towards AI agents. We use the results of this questionnaire as the baseline when we measure attitude towards AI *after* the experiment. (For the details of the questionnaire, see Sec. 3.2.

Phase 2: Human-AI mode The participant works *with* the AI agent. After a brief *Introduction of Phase 2*, the AI agent takes over the control over the game. Occasionally (roughly, every 50 steps), the AI agent will prompt the participant for advice, rendering this step a *collaborative game play*. The human participant responds by giving a directional advice using the directional buttons on the keyboard.

Different urgency modes. To allow us to answer RQ2 (the effect of urgency in interaction on trust), the participant performs two rounds of interactive game play. In one round (5 minutes), the participant has unlimited time to provide the advice. In the other round (5 minutes), the participant has to provide the advice in five seconds, otherwise the agent will disregard it (urgency).

Counterbalanced within-subject design. We use counterbalanced within-subject design [23], i.e, we expose participants to the two interaction modes in a different order. (Group 1: unlimited time interaction → limited-time interaction; Group 2: limited-time interaction → unlimited time interaction.)

After ten minutes of game play, the researcher ends the session, records the participant's score, and the participant fills in the questionnaire (Sec. 3.2).

[2] https://gymnasium.farama.org/v0.28.0/environments/atari/pacman/

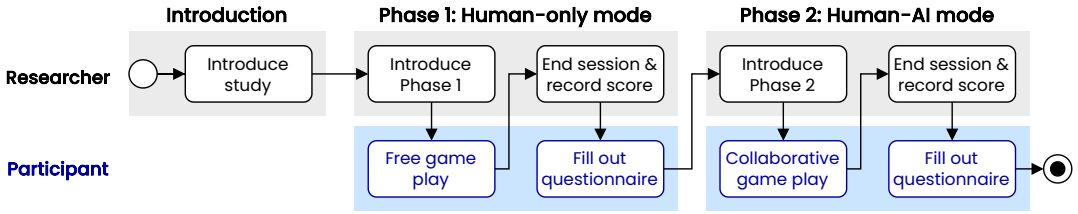


Figure 1: Experiment Overview

3.2 Questionnaire development

We collect data about demographics and trust attitudes towards AI agents. For the latter, we use Likert-type rating scales to guide participants in expressing their answers. Likert-type scales are psychometric rating scales often employed in questionnaires to measure the attitude of participants towards a specific statement [25]. Here, we measure the attitude of participants towards statements about trust and elements of trust in AI, as detailed in Sec. 2.1 and Tab. 1.

We design two questionnaires. The one filled in Phase 1 of the experiment (see Fig. 1) elicits the participant's trust attitude towards AI *in general*, before the interactive exercise. The one filled in Phase 2 elicits the participant's trust attitude specifically towards the AI agent in the experiment.

We use pairs of statements such as "*In general, I trust AI agents*" in the first questionnaire and "*I trust this particular AI agent*" in the second questionnaire. We ask similar pairs of questions about ease of use, dependability, predictability, alignment, and transparency—the applicable elements of trust extracted in Sec. 2.1 and Tab. 1. Our goal is to detect differences in the responses to these pairs before and after the intervention. To explain potential differences, in the second questionnaire, we elicit explanatory factors through Likert items such as "*Observing the AI agent in the experiment helped me understand the reasoning behind its decisions*" (RQ1), and Likert items related to time pressure (RQ2).

We use a five-point Likert scale of *1: Strongly disagree* to *5: Strongly agree*, and analyze the data accordingly.[3]

3.3 Design trade-offs and threats to validity

External validity. Our sample of experimental population may limit the generalization of results to a wider audience. To keep our experiment tractable, we recruited university students from computer science programs as participants. This population tends to be more experienced with AI tools than the general audience. However, this sample is still sufficient for our goal to measure *change* in attitudes in response to the interventions in our study.

Internal validity. Learning effect may give rise to internal validity. To mitigate this threat, we use a counterbalanced within-subject [23] design, i.e, expose participants to human-only and AI interaction modes in a different order and observe potential differences between the two groups. Slight threats arise from the overloaded notion of "*trust*." To mitigate this threat, we provided participants with definitions of potentially ambiguous concepts.

Construct Validity. In Likert-type data, a threat to construct validity stems from acquiescence bias, i.e., humans' tendency to agree with a statement in the questionnaire. Following questionnaire best practices, we mitigate this threat by using both positive and negative statements. Another threat to construct validity is social desirability, i.e., answering a questionnaire in a way participants think will make them look good. We mitigate this threat by making the questionnaire fully anonymized. Some threats to validity arise from the use of a Likert scales to assess perceived self-efficacy, instead of a better-suited psychometric scales, e.g., the generalized self-efficacy scale [51]. Opting for Likert scales was an early design choice when we decided to measure multiple factors, not only self-efficacy.

4 RQ1: Trust by human-AI interaction

To validate our study setup, we first corroborate that interactions with an AI indeed increase trust in our experiment, in line with observations from the state of the art [17, 62, 63, 65].

4.1 Improved trust in AI

Fig. 2 reports the attitude responses before and after the experiment. Shown in Fig. 2a, we measure a 50% increase in trust as the previous number of agreements increases from 20% to 70% (strong agreement and agreement in total). At the same time, disagreement decreases only slightly, from 30% to 24%. A fraction of people, 7%, remain neutral. A more detailed look in Fig. 3 reveals that the 50% post-intervention increase in trust is not solely due to previously neutral participants changing their attitude.

The eventual 70% of participants who express trust in the AI agent comprises 5 of 30 participants (16.7%) who were trustful before the intervention, 10 of 30 (33.3%) previously neutral, and 6 of 30 (20.0%) previously distrustful participants. Conversely, the previous 50% neutral stratum becomes more trustful in 10 of 30 cases (33.3%), remains neutral in 1 of 30 cases (3.3%), and becomes distrustful in 4 of 30 cases (13.3%).[4]

> ***Observation:*** *Trust increases substantially, especially due to both neutral and distrustful participants developing trust.*

As shown in Fig. 2b–2f, the known elements of trust do not change substantially, despite the clear increase in trust. The perceptions of most factors slightly improves. The only exception being ease of use (Fig. 2b), with a change of agreement from 90% to 87%. The number of participants who agree that the AI agent is dependable (Fig. 2c) increases by 6%, and the number of those who disagree

[3] The complete questionnaire is available in the replication package.

[4] More details are available in the replication package.

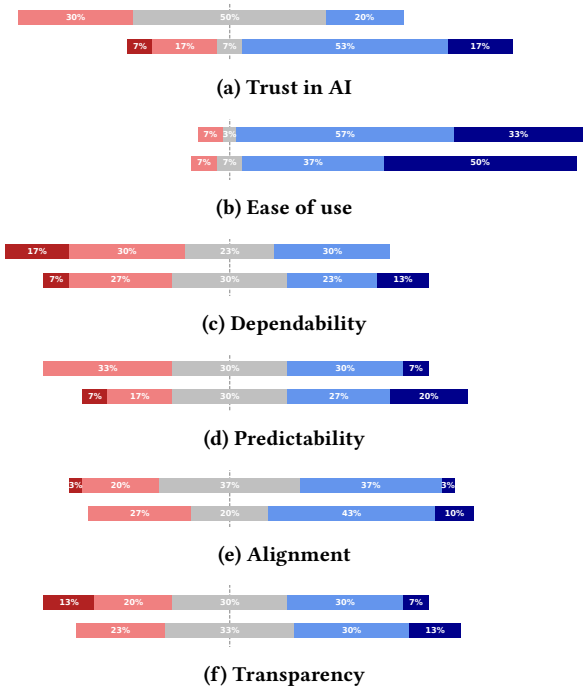


Figure 2: Participants' perception of trust in AI (*a*) and its elements (*b–f*) before (top) and after (bottom) a human-AI interactive experience

■ Strongly disagree ■ Disagree ■ Neutral ■ Agree ■ Strongly agree

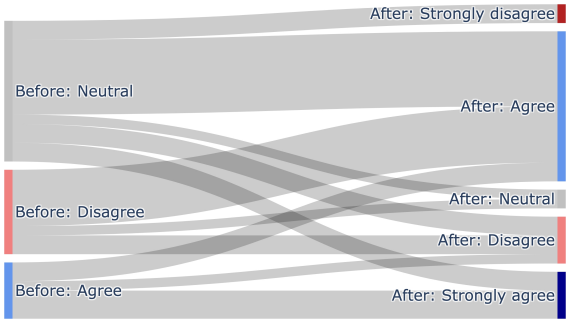


Figure 3: Changing trust attitudes of participants

with this assertion decreases by about 13%. A similar tendency can be observed in predictability (Fig. 2d) with an increase in agreement by 10% and a decrease in disagreement by 9%; in alignment (Fig. 2e) with an increase in agreement by 13% and a decrease in disagreement by 4%; and in transparency (Fig. 2f) with an increase in agreement by 6% and a decrease in disagreement by 10%.

> ***Observation:*** *Trust increases substantially, but known trust components remain essentially unaffected.*

Discussion. Our observations are in line with the established view in the state of the art, i.e., that trust increases with interaction. We did not detect change in the specific trust components that could pinpoint the key mechanism behind the improved trust attitude. One plausible explanation is that we did not select the right trust factors in our study design (Tab. 1). Another plausible explanation is that current trust taxonomies may not be able to describe causation between trust component and eventual trust. However, this investigation is not in the scope of our work; we merely aimed to establish that trust indeed increases in our setting.

4.2 Preference against AI autonomy

In the post-experiment questionnaire, we asked the participants for additional details, shown in Fig. 4, to be able to explain the mechanisms that effect trust.

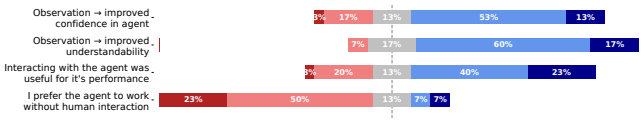


Figure 4: Perception of interaction with the AI agent. (Likert items reasonably paraphrased for brevity.)

■ Strongly disagree ■ Disagree ■ Neutral ■ Agree ■ Strongly agree

Following the interactive experience, 20 of 30 participants (66.7%) agree that their confidence in the agent improved by observing its behavior; and only 6 of 30 (20.0%) disagree with this assertion. 23 of 30 participants (76.7%) agree that the understandability of the agent improved by observing its behavior; and only 2 of 30 (6.7%) disagree. These figures demonstrate that observation alone may improve key attitudes towards the AI (confidence in, and understandability of AI). 20 of 30 participants (66.7%) also agree that the agent's performance improved by the active human-AI interaction.

23 of 30 participants (76.7%) disagree with the suggestion that the AI agent should work on its own. Thus, despite the improved confidence in and understandability of the AI agent, participants prefer to remain part of the collaboration. This observation excludes the possibility that humans' attitudes towards the AI agent changed positively due to the simplicity of the experimental task.

> ***Observation:*** *Human-AI interaction is a valued mechanism, and humans prefer not to allow the AI agent to work on its own.*

Discussion. Despite the improved trust towards the AI, the participants are still reluctant to allow more autonomy to the AI and let it operate without human interaction. This is an unexpected development considering that AI's superior ability to play simple video games is well-documented and well-known [46]. A plausible explanation is the lack of sufficient onboarding and therefore, the lack of willingness to give autonomy. Onboarding in AI software is often tied to the gradually increasing degree of autonomy the AI system is given, which increases trust calibration and usability compared to immediate full autonomy [29]. Another plausible explanation is that participants have pre-existing biases that our study was not designed to uncover.

RQ1: Improved trust by interaction

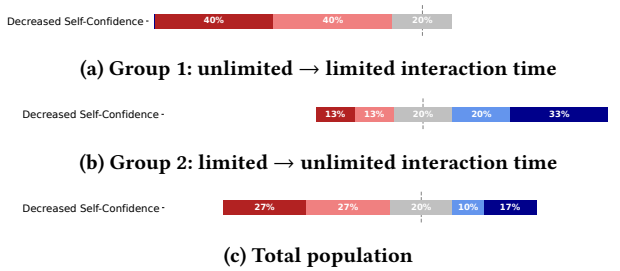
We corroborate that interactions with the AI agent indeed improve users' perceived trust in the agent. However, we remark that despite the increased trust, humans still prefer interactive human-AI collaboration rather than giving full autonomy to the AI.

5 RQ2: The impact of urgency

As explained in Sec. 3.1, in Phase 2, two groups of participants interacted with the AI differently. Participants in Group 1 first interacted with the AI agent without a time limit (i.e., no urgency) to provide advice, and subsequently, with time limit (i.e., urgency). Participants in Group 2 executed Phase 2 in the other way around, i.e., were exposed to urgency without prior experience with the AI.

5.1 Significant difference in self-confidence

The most important observation is shown in Fig. 5.



(a) Group 1: unlimited → limited interaction time

(b) Group 2: limited → unlimited interaction time

(c) Total population

Figure 5: Perceived effects of time pressure on self-confidence

■ Strongly disagree ■ Disagree ■ Neutral ■ Agree ■ Strongly agree

We observe substantial difference between the groups in terms of self-confidence. Most participants in Group 1, 12 of 15 (80.0%) disagree or strongly disagree that their self-confidence decreased; and no one agrees with this statement. Conversely, in Group 2, only 4 of 15 participants (26.7%) disagree and the majority of participants, 8 of 15 (53.3%) agree that their self-confidence has decreased.

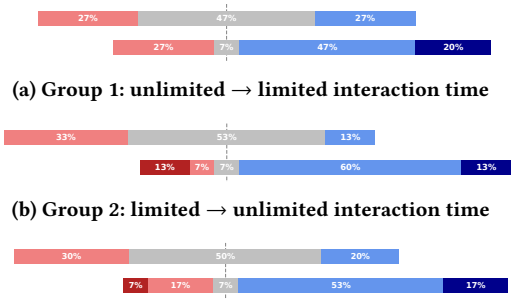
A chi-square test reveals that the difference between Group 1 and Group 2 is significant at $\alpha = 0.05$ with $p = 0.002$ ($\chi^2 = 12.00$), and the effect size (Cramér's V = 0.63) suggests strong association.

Observation: *Subjecting participants to time-pressured prompting by the AI has a statistically significant adverse impact on participants' self-confidence in our experiment.*

Discussion. These figures suggest that subjecting participants to time-pressured prompting by the AI may decrease participant's self-confidence and that, at significant levels. However, this effect disappears when participants are given the opportunity to first interact with the AI without time constraints, i.e., being eased into the collaboration with the AI. Indeed, it seems that with reasonable time to get familiar with an AI tool, participants feel more confident about their own work. As shown in Fig. 5c, this significant tendency of decreased self-confidence cannot be detected at the level of the total experimental population (i.e., by lumping the results from Group 1 and Group 2). It is, therefore, plausible that the difference between Group 1 and Group 2 in terms of self-confidence is due to the specific sequences of urgency modes.

5.2 No difference in trust and in the preference against AI autonomy

Fig. 6 shows that trust attitude towards the AI changes in the same way in the two groups, and this change is consistent with the change observed in the total population (previously reported in Sec. 4.1 and Fig. 2a). Distrust decreases and trust increases, eventually reaching about the same level of trust in both groups (Group 1: 10 of 15 participants (66.7%) agree or strongly agree; Group 2: 11 of 15 participants (73.3%) agree or strongly agree). This aligns with the trust attitude measured on the total population (21 of 30 – 70.0%), shown in Fig. 6c.



(a) Group 1: unlimited → limited interaction time

(b) Group 2: limited → unlimited interaction time

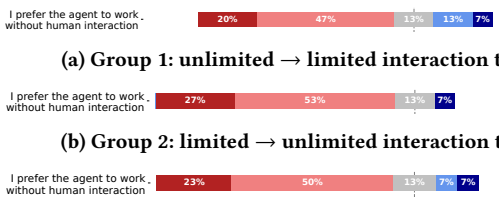
(c) Total population (as shown in Fig. 2a)

Figure 6: Attitude towards trust in the two groups before (top) and after (bottom) human-AI interaction

■ Strongly disagree ■ Disagree ■ Neutral ■ Agree ■ Strongly agree

Observation: *Urgency has no substantial effect on the eventual trust attitude. Trust increases similarly in both observed urgency modes.*

We do not observe a difference in participants' preferences for or against AI autonomy. As reported in Fig. 7, the two groups show a similar tendency towards *not* giving complete autonomy to the AI (Group 1: 10 of 15 participants (66.7%) disagree or strongly disagree; Group 2: 12 of 15 participants (80.0%) disagree or strongly disagree). This aligns with the preference measured on the total population (22 of 30 – 73.3%), shown in Fig. 7c.



(a) Group 1: unlimited → limited interaction time

(b) Group 2: limited → unlimited interaction time

(c) Total population (as shown in Fig. 4)

Figure 7: Perceived effects of time pressure on users' interaction preference (Group 1 and 2)

■ Strongly disagree ■ Disagree ■ Neutral ■ Agree ■ Strongly agree

> ***Observation:*** *Urgency has no substantial effect on the participants' preference against giving complete autonomy to the AI.*

Discussion. The loss of self-confidence in our experiments manifested in response to unprepared urgency. That is, Group 2 was not allowed an unlimited-time joint work with the AI agent before time pressure has been applied. Group 1 has had the opportunity to work with the AI agent without being rushed. We conjecture that this interaction allowed Group 1 to understand the impact of their own contributions to the joint problem-solving process and value themselves accordingly. Improved trust in the AI at the same time as loss of self-confidence is a cognitive asymmetry that may lead to cognitive dissonance which, in turn, may be detrimental to self-image, erode self-efficacy, and deteriorate self-esteem [56].

RQ2: The effects of (unprepared) urgency

Our results indicate that urgency during human-AI interaction may have substantial (even significant) impact on human users' self-confidence when humans are not given sufficient time to get familiar with the AI. At the same time, even when self-confidence drops, trust in the AI is maintained.

6 Discussion

We now discuss the results and elaborate on the key implications for software engineering researchers and practitioners.

6.1 On the cognitive asymmetry of reduced self-confidence and improved trust

One plausible explanation of the cognitive asymmetry between increased trust in the AI agent and decreased self-confidence is that introducing time pressure without proper training amplifies anxieties around AI's capabilities. Evidence from the state of the art shows that a common fear about AI—despite the obvious need for human intelligence and ingenuity in solving complex problems—stems from the reasoning capabilities of AI agents [18]. Humans may (and often do) attribute anthropomorphic properties to AI agents [14], and may feel pressured by the abilities of this perceived virtual "human" that exceed actual human skills. This, in turn, may lead to a loss of confidence [3, 15]. The increased focus we observe under time pressure (Fig. 5) may indicate a response to increased pressure in an attempt to keep pace with the AI. However, such cognitive pressure in long term is unlikely to be desirable as it is linked to increased burnout [48] and reduced productivity [5].

There are numerous human and technical consequences of our observations. It is easy to see how this cognitive asymmetry may lead to unsustainable human-AI collaboration on account of eroding self-efficacy, i.e., a person's belief in their ability to successfully organize and execute the actions required to achieve specific goals [10]. Low self-esteem and reduced perceived competence are strongly associated with anxiety and stress in longitudinal studies [54]. Evidence from automated decision support systems also shows that time pressure tends to reduce decision accuracy [45], further impacting humans' perceived self-efficacy. But anxiety and eroded self-esteem are not the only symptoms humans may develop in this setup. Humans may also develop overreliance on AI [60]. Although in our experiments, the participants overwhelmingly rejected the idea of allowing complete autonomy to the AI (Fig. 7), it is possible that repeated exposure to new AI tools and problems could shift the participants' attitudes. Such experiments are outside the scope of our study and are left for future work.

In general, our observations and the subsequent mechanisms can be framed by the theory of cognitive dissonance, the psychological discomfort that arises from holding conflicting cognitions [56]. Individuals have a motivation to seek consonance between their beliefs and actions [8], and if consonance cannot be established, individuals tend to change their beliefs rather than their actions [20], e.g., overestimating AI's indispensability or underestimating their own competence. It is plausible that unprepared urgency challenged the participants in our study in their belief that they are a useful party in the human-AI collaborative endeavor; to which the participants may have reacted by changing their belief, which manifested in the measured reduction in self-confidence.

From a technical point of view, these human implications may limit the automation level and adoption potential of AI. If AI tools stop being useful and empowering aids, their adoption will eventually slow down. It is, therefore, in every AI developer's and vendor's best interest to consider such barriers and actively help ease humans into human-AI settings.

6.2 Self-efficacy as a software quality attribute

To respond to the potential adverse implications of reduced self-confidence on self-efficacy—one's belief in their ability to successfully organize and execute the actions required to achieve specific goals [10]—we advocate for framing self-efficacy as a software quality attribute. Self-efficacy influences how people approach challenges and persist in challenging (professional) situations, e.g., the ones that require AI assistance. This makes a good case for asserting explicit value to self-efficacy in the design of AI-enabled systems.

Framing social and individual sustainability properties as a quality attribute of software is not a new idea. Such directions have been thoroughly explored in the related body of knowledge, notably in the seminal works of Lago et al. [31], Abrahão et al. [1], and Naveed et al. [39]. However, evidence shows that software engineering has limitations in embracing individual and social sustainability properties, [21, 53]. To address this shortcoming, individual and social sustainability properties and particularly, self-efficacy ought to be treated as a quality attribute that can be systematically assessed, optimized, and engineered into software systems.

Software engineering as a profession itself would benefit from such improvements as well. Self-efficacy in software engineering has been associated with satisfaction, performance, and engagement with work and their teams among software engineers in industry [43]. Software engineers with high self-efficacy exhibit more proactive, socially engaged, and confident behavior in development contexts [44]. Already at the early stage of engineering education, self-efficacy is a strong predictive marker of academic performance, motivation, persistence, and skill development [43]. Therefore, our recommendation to treat self-efficacy as a quality of software applies to software used by software engineers, too,

including IDEs, design tools, and documentation assistants. Modern IDEs already provide generative AI features, such as copilots, putting software engineers in a particularly susceptible position.

Promoting self-efficacy to a software quality attribute requires methodological and tool support, as well as thorough empirical evaluations. We recommend software engineering researchers to expand software evaluation methods by validated psychometric tools, standardized metrics, and experimental protocols, preferably of longitudinal nature to detect long-term tendencies of attitude change. Human factors research on trust and automation [22] provides an established foundation for incorporating self-efficacy–related measures into software engineering experiments. Research in human-computer interaction (HCI) and decision support indicates that transparent communication of uncertainty can improve calibrated trust and decision quality (e.g., visual and probabilistic uncertainty representations [27]; explanation and rationale-sharing [6]; and adaptive confidence displays [29]).

6.3 Organizational dimensions: socio-techno-economic risks and sustainable adoption

The implications of our observations extend beyond individuals' cognition and may impact the broader socio-techno-econmic context in which the investigated human-AI interactions are typical. Widespread digital transformation triggered rapid adoption of AI systems in organizations [55] with the expectation that AI will improve competitive advantage through increased productivity and problem-solving capabilities. However, if AI systems are not introduced and aligned with business processes properly, organizations may expose their employees to the unwanted effects outlined previously: lower self-confidence, decreased self-efficacy, less satisfaction from work, and anxiety. Time-pressure may also incentivize superficial decision-making [52], over-reliance on automation [58], or avoidance of responsibility [19]. These are just some of the socio-economic risks organizations face when engaging with AI systems.

Both from a social responsibility and economic point of view, organizations should feel motivated to develop mechanisms to mitigate these risks. Gradual onboarding and its variants—e.g., scaffolded introduction, progressive training—allow users to get familiar with the features and working modes of the system. Gradual onboarding is often tied to the increasing degree of autonomy an (AI) system is given, which increases trust calibration and usability compared to immediate full autonomy [29]. Our results, specifically in Fig. 7 show that full autonomy of AI is indeed not desired by the participants, even in a simple task as an arcade game (given that sufficient investment his facilitated). In other forms of onboarding, features are introduced gradually to users [7]. Such strategies could be used, e.g., to allow users to get comfortable with an agentic AI by restricting the AI's proactive prompting features, i.e., restricting it to a passive component the user controls. This removes urgency from the human-AI setting and invokes the mechanism we observed in experimental Group 1, i.e., self-confidence may be preserved. As demonstrated, onboarding is not a mere training activity but a complex deployment strategy organizations need to align with their digital transformation approach and existing business processes.

In companies with lower digital adeptness, upskilling plays an important role in establishing elementary working primitives that enable joint work between humans and AI [13]. Targeted training has also been shown to improve technology acceptance [37], reduce technology misuse [35], and increase performance through improved self-efficacy [59]. Specifically in the case of AI systems in a work environment, upskilling can help employees adopt accurate mental models of system capabilities and properly calibrate their self-confidence [28]. Organizations should be motivated to implement upskilling programs as such endeavors make organizations more resilient to future change—an added benefit that positively impacts the employees as well. Similar to onboarding, upskilling is not a simple training activity either, and must be situated in the complex socio-techno-economic landscape of organizations.

At the end of the day, sustainable adoption of AI systems demands that organizations consider the complex interplay of social, technical, and economic factors. In this work, specifically, we call for proper onboarding and upskilling mechanisms to retain human values and to facilitate employee satisfaction in an increasingly digitalized organization. Alas, social factors are still often overlooked as a sustainability factor [21]. This is apparent, e.g., in top-down strategic programs such as the Twin Transition, framed in the European Green Deal, that has a demonstrated blind spot for social and individual aspects in sustainable digital transformation [53].

We recommend researchers to investigate AI deployment strategies that regularly evaluate social and individual aspects in organizations, and align well with the organization's economic goals. Studying organizational adoption patterns and using structured frameworks, such as SusAF [11]

7 Conclusion

In this paper, we reported on our empirical study on how urgency in human-AI interaction may lead to reduced self-confidence. At the same time, trust in the AI agent typically still improves in humans. We draw the attention to this cognitive asymmetry because it has numerous potentially adverse implications on humans who interact with AI systems, e.g., reduced self-efficacy and performance. Given that the latest wave of mixed-initiative AI systems—e.g., agentic AI—often actively prompt humans, such urged interactions may appear at an increasing rate. We observe, however, that humans who are eased into a human-AI collaborative setting may overcome this problem and retain self-confidence. Thus, we recommend AI adopters (teams, organizations, companies) to consider upskilling and training before exposing humans to time-constrained situation.

We recommend researchers to investigate interaction mechanisms that aid the retention of self-confidence in human-AI settings, such as confidence calibration, and regular evaluation of psychometric properties. We advocate for promoting self-efficacy to a software quality metric to enable the systematic engineering of human-centered AI software, and we urge the development of supporting design methods, design principles, and UX patterns.

Acknowledgment. We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), DGECR2024-00293 (End-to-end Sustainable Systems Engineering).

References

[1] Silvia Abrahão, John Grundy, Mauro Pezzè, Margaret-Anne Storey, and Damian A. Tamburri. 2025. Software Engineering by and for Humans in an AI Era. *ACM Trans. Softw. Eng. Methodol.* 34, 5 (2025).

[2] Saleh Afroogh, Ali Akbari, Emmie Malone, Mohammadali Kargar, and Hananeh Alambeigi. 2024. Trust in AI: progress, challenges, and future directions. *Humanities and Social Sciences Communications* 11, 1 (2024), 1568.

[3] Amani Alabed, Ana Javornik, and Diana Gregory-Smith. 2022. AI anthropomorphism and its effect on users' self-congruence and self–AI integration: A theoretical framework and research agenda. *Technol. Forecast. Soc. Change* 182, 121786 (2022), 121786.

[4] James E Allen, Curry I Guinn, and Eric Horvtz. 1999. Mixed-initiative interaction. *IEEE Intelligent Systems and their Applications* 14, 5 (1999), 14–23.

[5] Shaimaa A. A. M. Amer, Sally Fawzy Elotla, Abeer Elsayed Ameen, Jaffer Shah, and Ahmed Mahmoud Fouad. 2022. Occupational Burnout and Productivity Loss: A Cross-Sectional Study Among Academic University Staff. *Frontiers in public health* 10 (2022).

[6] Saleema Amershi et al. 2019. Guidelines for Human-AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 1–13.

[7] Saleema Amershi et al. 2019. Software Engineering for Machine Learning: A Case Study. In *2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*. IEEE, 291–300.

[8] Elliot Aronson. 1969. *The Theory of Cognitive Dissonance: A Current Perspective*. Elsevier, 1–34.

[9] A. Avizienis, J.-C. Laprie, B. Randell, and C. Landwehr. 2004. Basic concepts and taxonomy of dependable and secure computing. *IEEE Trans Dependable Secure Comput* 1, 1 (2004), 11–33.

[10] Albert Bandura. 1977. Self-efficacy: Toward a unifying theory of behavioral change. *Psychol Rev* 84, 2 (1977), 191–215.

[11] Maike Basmer, Timo Kehrer, and Birgit Penzenstadler. 2021. SusAF Welcomes SusApp: Tool Support for the Sustainability Awareness Framework. In *2021 IEEE 29th International Requirements Engineering Conference (RE)*. 418–419.

[12] Meriem Ben Chaaben, Lola Burgueño, Istvan David, and Houari Sahraoui. 2026. On the Utility of Domain Modeling Assistance with Large Language Models. *ACM Trans. Softw. Eng. Methodol.* 35, 4, Article 112 (2026), 38 pages. doi:10.1145/3744920

[13] Tiziana C. Callari and Lucia Puppione. 2025. Meaningful work as shaped by employee work practices in human-AI collaborative environments: a qualitative exploration through ideal types. *Euro J of Inno Manag* 28, 10 (2025), 5001–5027.

[14] Xusen Cheng et al. 2022. Human vs. AI: Understanding the impact of anthropomorphism on consumer response to chatbots from the perspective of trust and relationship norms. *Inf. Process. Manag.* 59, 3 (2022), 102940.

[15] Leah Chong, Guanglu Zhang, Kosa Goucher-Lambert, Kenneth Kotovsky, and Jonathan Cagan. 2022. Human confidence in artificial intelligence and in themselves: The evolution and impact of confidence on adoption of AI advice. *Comp Hum Beh* 127, 107018 (2022), 107018.

[16] Hyesun Choung, Prabu David, and Arun Ross and. 2023. Trust in AI and Its Role in the Acceptance of AI Technologies. *Int J Hum Comput Interact* 39, 9 (2023), 1727–1739.

[17] Peter Denno. 2024. Cognitive work in future manufacturing systems: Human-centered AI for joint work with models. *J integ des&proc sci* 27, 2 (2024), 71–82.

[18] Christopher Diebel, Marc Goutier, Martin Adam, and Alexander Benlian. 2025. When AI-based agents are proactive: Implications for competence and system satisfaction in human–AI collaboration. *Bus. Inf. Syst. Eng.* (2025).

[19] Marwa El Zein, Bahador Bahrami, and Ralph Hertwig. 2019. Shared responsibility in collective decisions. *Nature Human Behaviour* 3, 6 (2019), 554–559.

[20] Leon Festinger. 1957. *A Theory of Cognitive Dissonance*. Stanford University Press, Palo Alto, CA.

[21] Johanna Liz Gustavsson and Birgit Penzenstadler. 2020. Blinded by Simplicity: Locating the Social Dimension in Software Development Process Literature. In *Proc of the 7th Intl Conference on ICT for Sustainability*. ACM, 116–127.

[22] KA Hoff and Masooda Bashir. 2015. Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust. *Hum Fac* 57, 3 (2015), 407–434.

[23] Rajiv Jhangiani, I-Chant A Chiang, Carrie Cuttler, and Dana C Leighton. 2019. *Research methods in psychology*.

[24] Jiaming Ji et al. 2025. AI Alignment: A Comprehensive Survey. arXiv:2310.19852 [cs.AI]

[25] Ankur Joshi et al. 2015. Likert scale: Explored and explained. *British Journal of Applied Science & Technology* 7, 4 (2015), 396.

[26] Vanessa Juth et al. 2008. How Do You Feel?: Self-esteem Predicts Affect, Stress, Social Interaction, and Symptom Severity during Daily Life in Patients with Chronic Illness. *J Health Psychol* 13, 7 (2008), 884–894.

[27] Matthew Kay, Tara Kola, Jessica R. Hullman, and Sean A. Munson. 2016. When (ish) is My Bus? User-centered Visualizations of Uncertainty in Everyday, Mobile Predictive Systems. In *Proc 2016 CHI Conf on Human Factors in Computing Systems (CHI '16)*. ACM, 5092–5103.

[28] Suhas S Khot and Neha N Ganvir. 2024. TECHNOLOGY-ENABLED AND MOBILE-ORIENTED WORLD WITH SMART PRODUCTS: A MULTIPLE HOLISTIC APPROACH. *Weser Books* (2024), 1.

[29] Rafal Kocielnik, Saleema Amershi, and Paul N. Bennett. 2019. Will You Accept an Imperfect AI? Exploring Designs for Adjusting End-user Expectations of AI Systems. In *Proc 2019 CHI Conf on Human Factors in Comp Sys*. ACM, 1–14.

[30] Nagendra Kumar, Rajeev Ranjan Kumar, and Alok Raj and. 2025. Establishing Antecedents and Outcomes of Human-AI Collaboration: Meta-Analysis. *Journal of Computer Information Systems* (2025), 1–15.

[31] Patricia Lago, Sedef Akinli Koçak, Ivica Crnkovic, and Birgit Penzenstadler. 2015. Framing sustainability as a property of software quality.

[32] Jingshu Li, Yitian Yang, Q. Vera Liao, Junti Zhang, and Yi-Chieh Lee. 2025. As Confidence Aligns: Understanding the Effect of AI Confidence on Human Self-confidence in Human-AI Decision Making. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. ACM, Article 1111.

[33] Xiaoran Liu and Istvan David. 2026. A Reference Architecture of Reinforcement Learning Frameworks. arXiv:2603.06413 [cs.SE] https://arxiv.org/abs/2603.06413

[34] Shuai Ma, Xinru Wang, Ying Lei, Chuhan Shi, Ming Yin, and Xiaojuan Ma. 2024. "Are you really sure?" Understanding the effects of human self-confidence calibration in AI-assisted decision making. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–20.

[35] Janet H. Marler, Xiaoya Liang, and James Hamilton Dulebohn. 2006. Training and Effective Employee Information Technology Use. *Journal of management* 32, 5 (2006), 721–743.

[36] Roger C. Mayer, James H. Davis, and F. David Schoorman. 1995. An Integrative Model Of Organizational Trust. *Acad Manage Rev* 20, 3 (1995), 709–734.

[37] Monica Molino, Claudio G. Cortese, and Chiara Ghislieri. 2020. The Promotion of Technology Acceptance and Work Engagement in Industry 4.0: From Personal Resources to Information and Training. *Int J Env Res and Pub Health* 17, 7 (2020).

[38] Ismail Mseer, Khaled Al-Qawasmi, Mansoor Alaali, Arafat Salih Aydiner, and Abdalmuttaleb M. A Musleh Al-Sartawi. 2025. Autonomous Artificial Intelligence: Revolutionizing the World. In *The Paradigm Shift from a Linear Economy to a Smart Circular Economy*. Vol. 586. Springer, 1319–1327.

[39] Hira Naveed, John Grundy, Chetan Arora, Hourieh Khalajzadeh, and Omar Haggag. 2024. Towards Runtime Monitoring for Responsible Machine Learning using Model-driven Engineering. In *Proceedings of the ACM/IEEE 27th International Conference on Model Driven Engineering Languages and Systems*. ACM, 195–202.

[40] Ume Nisa, Muhammad Shirazi, Mohamed Ali Saip, and Muhammad Syafiq Mohd Pozi. 2025. Agentic AI: The age of reasoning—A review. *J Aut and Intell* (2025).

[41] Cecilie Bertinussen Nordheim, Asbjørn Følstad, and Cato Alexander Bjørkli. 2019. An Initial Model of Trust in Chatbots for Customer Service—Findings from a Questionnaire Study. *Interacting with computers* 31, 3 (2019), 317–335.

[42] Yeon Su Park, Nadia Azzahra Putri Arvi, Seoyoung Kim, and Juho Kim. 2026. Authorship Drift: How Self-Efficacy and Trust Evolve During LLM-Assisted Writing. *arXiv preprint arXiv:2602.05819* (2026).

[43] Jason Richard Power, David Tanner, and Jeffrey Buckley. 2024. Self-efficacy development in undergraduate engineering education. *European Journal of Engineering Education* (July 2024), 1–25.

[44] Danilo Ribeiro et al. 2023. Understanding Self-Efficacy in Software Engineering Industry: An Interview study. In *Proceedings of the 27th International Conference on Evaluation and Assessment in Software Engineering (EASE '23)*. ACM, 101–110.

[45] Tobias Rieger and Dietrich Manzey. 2022. Human performance consequences of automated decision aids: The impact of time pressure. *Human factors* 64, 4 (2022), 617–634.

[46] Sebastian Risi and Mike Preuss. 2020. From chess and Atari to StarCraft and beyond: How game AI is driving the world of AI. *KI - Künstl. Intell.* 34, 1 (2020), 7–17.

[47] Philipp Rohlfshagen, Jialin Liu, Diego Perez-Liebana, and Simon M. Lucas. 2018. Pac-Man Conquers Academia: Two Decades of Research Using a Classic Arcade Game. *IEEE Trans Games* 10, 3 (2018), 233–256.

[48] Wilmar B. Schaufeli and Toon W. Taris. 2014. A Critical Review of the Job Demands-Resources Model: Implications for Improving Work and Health. In *Bridging Occupational, Organizational and Public Health*. Springer, 43–68.

[49] Silvia Schiaffino and Analıá Amandi. 2004. User – interface agent interaction: personalization issues. *Int J Hum Comput Stud* 60, 1 (2004), 129–148.

[50] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. Proximal Policy Optimization Algorithms. arXiv:1707.06347 [cs.LG]

[51] Ralf Schwarzer and Matthias Jerusalem. 1995. Generalized self-efficacy scale. *Measures in health psychology: A user's portfolio. Causal and control beliefs* 35, 37 (1995), 82–003.

[52] Lennart Seitz, Sigrid Bekmeier-Feuerhahn, and Krutika Gohil. 2022. Can we trust a chatbot like a physician? A qualitative study on understanding the emergence of trust toward diagnostic chatbots. *Int J Hum Comput Stud* 165 (2022), 102848.

[53] Baran Shajari and Istvan David. 2025. Bridging the Silos of Digitalization and Sustainability by Twin Transition: A Multivocal Literature Review. In *2025 11th Intl Conf on ICT for Sustainability (ICT4S)*. 23–34.

[54] Julia Friederike Sowislo and Ulrich Orth. 2013. Does Low Self-Esteem Predict Depression and Anxiety? A Meta-Analysis of Longitudinal Studies. *Psychological*

bulletin 139, 1 (2013), 213–240.

[55] Mariagrazia Squicciarini and Heike Nachtigall. 2021. Demand for AI skills in jobs: Evidence from online job postings. *OECD Science, Technology and Industry Working Papers* 2021, 3 (2021), 1–74.

[56] Jeff Stone and Joel Cooper. 2001. A Self-Standards Model of Cognitive Dissonance. *J Exp Soc Psychol* 37, 3 (May 2001), 228–243.

[57] Virginia Sullivan and Kristin Weger. 2025. Transparency and Explainability in AI-Assisted Decision Making: Effects on Trust, Perceived Reliability, Confidence, and Ease of Understanding. *Proc Hum Fact and Ergo Soc Annual Meeting* (2025).

[58] Monica Tatasciore and Shayne Loft. 2024. Can increased automation transparency mitigate the effects of time pressure on automation use? *Applied Ergonomics* 114 (2024), 104142.

[59] Reza Torkzadeh, Kurt Pflughoeft, and Laura Hall. 1999. Computer self-efficacy, training effectiveness and user attitudes: An empirical study. *Behaviour & information technology* 18, 4 (1999), 299–309.

[60] Daniel Ullrich, Andreas Butz, and Sarah Diefenbach. 2021. The Development of Overtrust: An Empirical Simulation and Psychological Analysis in the Context of Human–Robot Interaction. *Frontiers in robotics and AI* 8 (2021), 554578–.

[61] Warren J von Eschenbach. 2021. Transparency and the black box problem: Why we do not trust AI. *Philos. Technol.* 34, 4 (Dec. 2021), 1607–1622.

[62] Qian Wan et al. 2024. "It Felt Like Having a Second Mind": Investigating Human-AI Co-creativity in Prewriting with Large Language Models. *Proc ACM Hum Comput Interact* 8, CSCW1 (2024).

[63] Jingda Wu, Zhiyu Huang, Zhongxu Hu, and Chen Lv. 2023. Toward Human-in-the-Loop AI: Enhancing Deep Reinforcement Learning via Real-Time Human Guidance for Autonomous Driving. *Engineering* 21 (2023), 75–91.

[64] Rongbin Yang and Santoso Wibowo. 2022. User trust in artificial intelligence: A comprehensive conceptual framework. *Electron Mark* 32, 4 (2022), 2053–2077.

[65] Guanglu Zhang, Leah Chong, Kenneth Kotovsky, and Jonathan Cagan. 2023. Trust in an AI versus a Human teammate: The effects of teammate identity and performance on Human-AI cooperation. *Comp Hum Behav* 139 (2023), 107536.

[66] Jianlong Zhou et al. 2020. Effects of personality traits on user trust in human–machine collaborations. *J multimodal user interfaces* 14, 4 (2020), 387–400.