Co-intelligence in design: the importance of trust in artificial intelligence

Chiara Lelli [1,2], **Filippo Chiarello** [1,2,✉] **and Vito Giordano** [1,2]

[1] *Università di Pisa, Italy,* [2] *Business Engineering for Data Science Lab, Italy*

✉ filippochiarello.90@gmail.com

ABSTRACT: As Generative Artificial Intelligence (GenAI) gets integrated in design processes, building trust in these systems is critical for effective human-AI collaboration. This study introduces a framework aimed at translating the abstract concept of trust into practical strategies for design teams, focusing on four trust factors: transparency, accountability, similarity, and performance. We tested the framework's impact on trust-building and trust learning using a mixed-methods approach, incorporating design tasks and structured workshops involving university students. The results highlight the framework's ability to enhance participants' understanding of trust in AI. Insights from this study contribute to advancing educational approaches for embedding trust in AI-driven design, revealing that design activities alone are not enough to impact trust learning.

KEYWORDS: trust, artificial intelligence, design education, human-AI collaboration, education

1. Introduction

As Generative AI (genAI) becomes more important, traditional design team dynamics evolve. These hybrid teams lead to performance levels that surpass what either could achieve alone (Mollick, 2024). In particular, Large Language Models (LLMs) have been applied in multiple areas of design, generating design ideas, improving communication within design teams (Chiarello et al., 2024) and producing training materials for design courses (Meron & Araci, 2023). A key factor in the successful adoption of genAI is its acceptance by the users, who need to interact with it and integrate their behaviors with AI systems (Glikson & Woolley, 2020). Analyzing acceptance models applications, Kelly et al., 2022 found that trust is a crucial factor influencing Artificial Intelligence (AI) acceptance. Trust influences performance, as different levels of trust can lead to disuse (rejection), misuse (overreliance) and abuse (harmful use) of them (Glikson & Woolley, 2020). Recent research in Engineering Design (ED) has addressed trust in relation to technologies like Digital Twins, Automated Vehicles, and AI systems, focusing on transparency, explainability and system-level enablers. Despite this attention, existing literature primarily addresses trust during AI system design, leaving a gap in understanding how trust evolves during their use.

To overcome this issue, in this article we present a framework to support designers in (1) translating the abstract and intangible concept of trust into actionable advice, and (2) to educate novice designers on the theme of trust. We tested the framework within a conceptual design project focusing on the early-stage ideation phase. Its efficacy was evaluated using statistical methods, on data collected through three questionnaires repeated in three key moments of the study. The purpose of this study was to investigate the following research questions:

- *RQ1*: Does the use of the framework impact *Trust Learning*?
- *RQ2*: Does the use of the framework impact trust? If so, which construct of the ones addressed in the German Technology in Automation (TiA) Questionnaire (Körber, 2019)?

Trust Learning refers to the level of knowledge participants have on the topic of trust in LLMs. It includes concepts and definitions, but also the ability to assess the LLMs trustworthiness and suggest improvements to enhance it. We prove that the use of the framework influences *Trust Learning*. The experiment also has a correlation with an increased participants *Familiarity* with Generative LLMs and reduced their perceived *Understandability*.

2. Trust foundations and levers in AI

The study of trust in AI spans various fields. Researchers in areas such as Computer Science and Psychology have identified key factors that influence people's trust in AI. In this section we review the main works on the topic, and we link this literature to the discourse on the importance of trust in ED.

2.1. Trust in AI

The concept of trust refers to "a psychological state comprising the intention to accept vulnerability based on positive expectations of the intentions or behaviors of another" (Rousseau et al., 1998). Traditional models of trust focus on interpersonal relationships between humans such as Mayer et al. 1995. However, for AI systems, models of trust in technology (McKnight et al., 2011) and automation (Lee and See, 2004) are also relevant. Analyzing the recent studies in the field of trust in AI, four factors were identified to be foundational: *transparency, accountability, similarity, and performance*.

Transparency refers to "the extent to which the operating rules and inner workings of the technology are visible to users" (Glikson & Woolley, 2020). Effective transparency involves providing clear and understandable explanations, which fosters trust in the AI system. While a lack of explanation can lead to distrust, overly detailed or complex explanations may overwhelm users and cause confusion.

Accountability is described as "the obligation to report and justify one's actions to an authority" (Novelli et al., 2023). For AI, this means defining the relationship between the AI system and its users, including how tasks are delegated, assessed, and the consequences of these assessments. Accountability involves not only answering for actions but also ensuring that the AI system's conduct meets established standards and procedures.

The *similarity* factor reflects the degree to which people's mental models align with that of the AI. In human relationships, finding common values or beliefs, enhances trust and fosters stronger connections. When it comes to AI, similarity is evaluated based on how closely the AI resembles humans, physically and intellectually, in the process and in the output, and how well its values and goals align with those of the team. A reasonable level of similarity makes the AI system feel more relatable and reliable, boosting trust. However, if the AI's resemblance becomes too pronounced it may provoke discomfort or mistrust, an effect known as Uncanny Valley (Troshani et al., 2021).

Performance refers to "the competency or expertise of the automation as demonstrated by its ability to achieve the operator's goals" (Lee & See, 2004). This factor evaluates what an AI system can accomplish and how reliably it performs these tasks. Performance includes the AI system's functionality, its consistency in delivering results, and the fairness of its outputs, as users are more inclined to trust an AI system that reliably meets their objectives and performs well in relevant tasks.

2.2. The importance of human-AI trust in design

The importance of trust in ED has already been acknowledged. Wijngaards et al., 2004 analyzed the importance of trust in distributed design processes, showing how trust thresholds influence decisions, such as task delegation, and affect design quality and team efficiency. Although they do not address AI in their assessment, it still suggests the importance trust has in design processes. Considering the increasingly active role AI assumes as a team member within human-AI teams (Larson & DeChurch, 2020), the importance of trust becomes even more critical. In more recent years, trust has been addressed when treating technologies such as Digital Twins (Trauer et al., 2022) and Automated Vehicles (She et al., 2021). In other cases, such as Clarke et al., 2021, the focus has been on interpersonal trust, mainly in the field of co-design. When handling trust in AI, the goal in ED has been to enhance trust towards AI systems leveraging on multiple aspects: expanding the function failure modes taxonomy for intelligent systems with embedded AI components (Campean et al., 2024), creating systems to increase the AI transparency and explainability (Hu et al., 2024) and implementing trust enablers to be embedded in the design of AI systems (Song et al., 2024).

Although a strong focus on trust exists, it primarily centers on its implementation during the design stages of AI systems, leaving a gap in addressing how trust can be actively influenced when using these systems.

3. The AI trust framework

In this section we present the AI Trust Framework. As shown in Figure 1, the framework is organized as a table with trust *factors* listed as rows and *levers* as columns. The framework is employed when AI is first integrated in a design team workflow. It is firstly introduced by a lecture on trust that provides basic insights on trust and the framework logic. Using the framework before, during or after the project, individually or in groups, the design team can reflect on their level of trust towards AI and identify strategies to face possible obstacles.

The selection of the *factors* (that are the same reviewed in section 2.1) was driven by prioritizing clarity in communication over strict scientific rigor. The proposed framework is not intended to introduce a new model of trust but rather to identify four components that are easily comprehensible to individuals without prior expertise in the field of trust, as we expect in the context of ED.

The *levers* follow Ishikawa's 5M model (Liliana, 2016), specifically adapted to the context of AI implementation. By considering trust factors as potential aspects where the AI system might fail, the analysis of the 5M framework, with the addition of an Environment factor, can help indicate the pain points and gain points of the project.

The framework operates under the assumption that the four trust factors can no longer be altered in the selected AI systems (costs and time to modify them are outside the scope of the project). These factors can create an obstacle to the acceptance of AI, as well as facilitate it. By creating a discussion on the *factors* influencing AI deployment through the lens of trust, the framework builds awareness and facilitates dialogue and collaboration in the design team. It also helps design teams identify the levers for AI introduction tailored to the specific project and the people involved.

The user of the framework fills in the matrix with green or orange notes. Green notes indicate factors that positively support the deployment and use of AI. On the contrary, orange notes show obstacles and barriers that hinder AI implementation and acceptance. Finally, the user of the framework can fill the last row of the matrix, labeled *Outcomes*. This row consolidates the ideas from each column, outlining actionable steps to enhance the AI introduction process.

Managerial Levers / Trust in AI Factors	Men	Machines	Materials	Methods	Measures	Environment
Transparency	Having a solid knowledge about AI allowed us to make a precise prompt			gpt gave us an explaination of how he got to a certain solution when asked to		we had the fear of getting results too similar to others
Accountability		gpt wasn't capable of understanding and correcting errors	the data given to the model wasn't accurately checked before sumbitting the promt		we knew that there was a uncertainty in the outcome	
Similarity	getting the desired results required submitting more promts					
Performance	most of the trust in the results was because of the lack of time	faster in evaluating ideas based on a context		we gave the model a precise context in order to get the result in the form we wanted	the results were good enough in a small amount of time	seeing how other people used AI model gave us an idea of it's behavior and how to get better results
Outcomes	learning more concepts about how AI works and organize the work could improve our outcomes		the data used should be revised first and then sumbitted to have more clarity with the results			we should improve the clarity and uniqueness of our prompts

Figure 1. Example of completed trust framework

3.1. Brainstorming vs questionnaire

The framework can be completed in two ways: through brainstorming sessions or questionnaires. The **brainstorming approach** stimulates discussion among groups of participants, allowing for diverse perspectives and comprehensive team awareness. The facilitator's role is vital in ensuring effective brainstorming (Oxley et al., 1996). They must communicate the framework's purpose and structure clearly, adapting their language to suit participants' diverse backgrounds and knowledge. In contrast, the **questionnaire approach** allows individuals to complete the framework without the pressure of group dynamics, enabling honest feedback. This method allows for simultaneous participation by a larger number of people, facilitating extensive data collection. The choice between brainstorming and questionnaires depends on the specific goals of the framework and the desired balance between interactive discussion and extensive data collection.

3.2. Key roles for framework implementation

It is crucial to carefully select who will work on completing the framework, to offer different viewpoints and promote homogeneous team awareness. Five main project roles have been identified:

- *Sponsor*: The individual or organization promoting the AI introduction. Their involvement helps them understand the potential barriers they may face.
- *Execution Team*: The group directly interacting with the AI, serving as intermediaries between the AI system and the organization. Their insights are crucial for understanding the practical aspects of AI integration.
- *Representative*: Individual business units, or functions with designated AI product owners and business analysts. The Spoke oversees execution teams, ensures solution adoption, and tracks performance.
- *Hub*: A central team responsible for overseeing the process, providing guidance, and evaluating the *Representative*'s decisions.
- *External Experts*: Specialists with experience in AI deployment and use. Their external perspective is especially valuable for proactive use of the framework, providing insights that might not be apparent to internal stakeholders.

3.3. Proactive vs reactive approach

The framework can be employed in two key scenarios: a reactive approach or a proactive one. In a **reactive approach**, it helps evaluate the current state of the AI deployment process, assessing how a range of factors impact trust and acceptance. After project completion, it highlights best practices and areas for improvement, aiding in the development of guidelines for future initiatives. In a **proactive approach**, the framework is used before the project begins to assess and optimize the initial trust situation, setting the stage for a more favorable deployment.

4. Testing of the AI trust framework

We conducted a user study on 30 undergraduate design students from the University of Pisa, Italy. The experiment, as shown in Figure 2, consisted of two main components: a design activity and a subsequent lecture-workshop session on the trust framework.

4.1. Design activity

In the design activity students were required to use ChatGPT-4o mini to generate and then identify three optimal solutions to a given need. The design activity provided a shared experience that students could reference while engaging with the trust framework. Despite all participants declaring prior experience with generative LLMs like ChatGPT, the activity ensured a consistent context using a specific AI model, ChatGPT-4o mini.

The design activity comprised two phases, divergent and convergent, and simulated project roles and collaborative dynamics. Students were divided into groups of three or four members. Each group appointed a *Representative* responsible for coordination and communication with the organizers but restricted from using ChatGPT-4o mini. Other group members formed the *Execution Team*, tasked with generating and evaluating solutions. This division reflected two of the key roles for framework
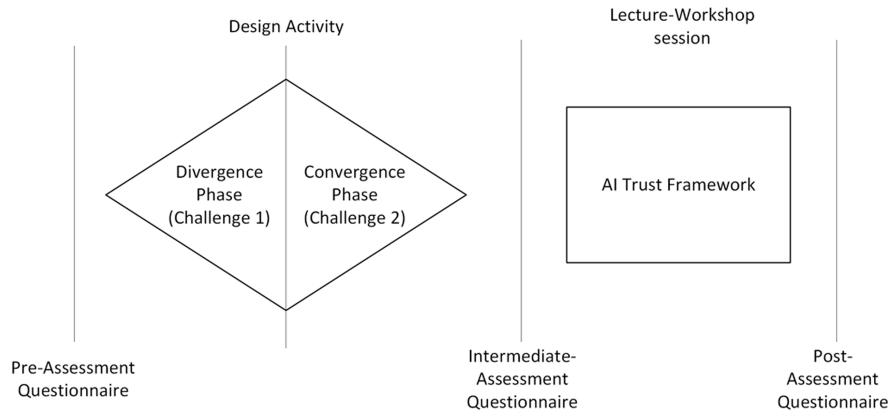
Figure 2. Experiment structure

implementation previously identified, while the other roles were excluded due to participant and time constraints. In the **divergent phase**, students generated as many solutions as possible to address the given need. In the **convergent phase**, they selected the three best solutions from those generated earlier. Both phases used the **Input-Process-Output (IPO) structure** to describe the solutions, with word limits of 30 and 120 words for the divergent and convergent phases, respectively. **Evaluation criteria** were adapted from the Torrance Tests of Creative Thinking (TTCT) (Torrance, 1966): fluency (quantity of solutions), flexibility (diversity of solutions), originality (novelty of solutions), and an added measure of quality (adherence to the IPO structure and word limits). These metrics were inspired by prior studies (e.g., Paschen et al., 2019; Urban et al., 2024), with "elaboration" excluded due to the imposed word limits. For the convergent phase, evaluation metrics aligned with industry standards for idea and concept screening (Hart et al., 2003): technical feasibility, product uniqueness, and "fit with the need" (substituting market potential for relevance to student understanding).

4.2. Lecture-workshop session

The second section comprised a lecture and a workshop on the AI Trust Framework. Data were collected using three questionnaires administered at three points: prior to the design activity, between the design activity and the lecture-workshop session and following the lecture-workshop session.

A lecture introduced students to basic concepts of trust in AI and described the trust framework. Groups were then given 40 minutes to collaboratively complete the framework, employing a **reactive brainstorming approach**. They reflected on the design activity, identifying factors that influenced their trust in the AI system, both as enablers and obstacles.

4.3. Questionnaire structure

To evaluate the impact of the experiment, three questionnaires, **Pre-Assessment**, **Intermediate-Assessment**, and **Post-Assessment**, were administered. The moments in which each questionnaire was implemented is shown in figure 2. These questionnaires repeated three of the five sections from the Pre-Assessment Questionnaire to track changes in key constructs over time.

The **Pre-Assessment Questionnaire** was divided into five sections: *Anagraphic Questions, Use of AI, Trust Factors, Trust learning, and Trust in Generative LLMs*. For the last three sections, participants rated their agreement with each statement on a Likert scale from 1 (Strongly Disagree) to 5 (Strongly Agree). Below is a detailed breakdown of each section:

1) *Anagraphic Questions*: Participants provided details on age, gender, nationality, educational level, and prior studies.
2) *Use of AI*: This section gathered data on participants' prior experience with generative LLMs. Questions focused on the frequency of use, application areas, and the specific generative LLMs employed.
3) *Trust Factors*: The goal of this section was to measure the importance participants attributed to the factors in the trust framework and how this importance evolved with experience (design activity) and knowledge (lecture and workshop). Each item corresponded to one trust factor

(transparency, accountability, similarity and performance) or to one lever (men, materials, machines, methods, measures and environment) in the framework.

4) *Trust Learning*: This section evaluated participants' learning of trust concepts. Items were based on the revised Bloom's taxonomy, thus representing the learning outcomes of the lecture-workshop session. Each item corresponded to a level in Bloom's taxonomy (remembering, understanding, applying, analyzing, evaluating, and creating), expressed as a self-assessment statement, from "I know basic concepts related to trust in Generative LLMs systems." (remembering) to "I can suggest improvements or guidelines to enhance the trustworthiness of Generative LLM systems." (creating).

5) *Trust in Generative LLMs*: This section measured participants' trust in generative LLMs, specifically ChatGPT-4o mini. Following existing literature reviews on trust questionnaires (Razin & Feigh, 2023), Körber's German TiA (Körber, 2019) was used as it showed high values of validity and reliability. Körber's questionnaire consists of six subscales: Reliability, Understandability, Propensity to Trust, Intention of Developers, Familiarity, and Trust in Automation. Each construct is strongly aligned with the factors presented in the framework, except for the factor of similarity. This instrument was preferred over others, such as McKnight's survey (McKnight et al., 2011), due to its concise format, which minimized respondent burden. Minor modifications were made to fit the experiment's context.

The **Intermediate**- and **Post-Assessment Questionnaires** retained three of the five sections from the Pre-Assessment Questionnaire: *Trust Factors, Trust Learning*, and *Trust in Generative LLMs*. This enabled the tracking of changes in these constructs after the design activity and framework lecture-workshop. Additionally, they recorded participants' roles (Representative or Execution Team member) to facilitate future analyses of role-specific variations.

5. Results and discussion

5.1. Sample characteristics

We collected data from 30 students with ages ranging from 20 to 28 who participated in all the phases of the experiment. Most of the participants hold a bachelor's degree. The sample exhibits a diverse distribution, including participants from various countries and backgrounds, ranging from Computer Science to Humanities. This diverse mix of educational backgrounds was selected to reflect the interdisciplinarity of real-world design teams, where professionals from various fields collaborate to solve complex problems. Nonetheless, all participants were enrolled in a Design course. Table 1 reports sample characteristics as collected from the *Anagraphic Questions* section of the questionnaire.

Table 1. Sample characteristics.

Variable	Value	Frequents	% Participants
Gender	*Female*	14	46.6
	Male	16	53.4
Nationality	*Italian*	13	43.4
	Indian	3	10
	Indonesian	3	10
	Spanish	3	10
	Polish	2	6.6
	Vietnamese	2	6.6
	*Other**	4	13.3
Educational Level	*High School*	4	13.3
	Bachelor's degree	22	73.4
	Master's degree	4	13.3

(*Continued*)

Table 1. Continued.

Variable	Value	Frequents	% Participants
Educational field	*Computer Science/Information Technology*	13	43.4
	Engineering	4	13.3
	Business/Economics	4	13.3
	Humanities (e.g. History, Languages, Philosophy)	3	10
	Mathematics/Statistics	3	10
	*Other***	3	10

*includes Ethiopian, Mongolian, Ossetian and Somali.

**includes Digital Humanities and Physics

All participants had previous experience with generative LLMs and reported using them at least monthly (Table 2). The main areas of use were university projects and assignments (87%), academic studies such as research assistance and Q&A (83%) and personal projects like coding, art and content creation (60%). The most used LLMs were ChatGPT (100%), Gemini (33%) and Claude (10%).

Table 2. LLM use in Sample.

Variable	Value	Frequents	% Participants
LLM use	*More than once a day*	8	26.7
	Daily	15	50
	Weekly	6	20
	Monthly	1	0.3

5.2. Descriptive analysis

As indicated by Körber, before starting the analysis, the responses to inverted items were recoded so that higher agreement resulted in lower scores in the corresponding construct. We then conducted a descriptive analysis on the gathered data on each construct in Körber's questionnaire and on the learning items of the Trust Learning section. The results of the three questionnaires were then compared to analyze possible variations. Figure 3 shows the boxplots of the distribution of mean values across the six constructs evaluated in the questionnaire (*Learning, Familiarity, Intention of Developers, Propensity to Trust, Reliability, and Understandability*) measured at three distinct stages: **Pre-Assessment** (blue), **Intermediate-Assessment** (orange), and **Post-Assessment** (pink). The y-axis represents the mean values, ranging from 1 to 5 (the range of the Likert scale), while the x-axis displays the evaluated factors. Each boxplot depicts the interquartile range (IQR), with the central line indicating the median, the upper and lower boundaries of the box representing the 75th and 25th percentiles, and the whiskers extending to the smallest and largest values within 1.5 times the IQR. Outliers beyond this range are marked as individual points.

From figure 3, it is qualitatively evident that the *Intention of Developers, Propensity to Trust* and *Reliability* constructs remain unchanged throughout the experiment. On the contrary, variations are present in the *Familiarity* and *Understandability* constructs from Körber's test. Moreover, the *Learning* construct based on Bloom's Taxonomy shows an evident variation. Next section provides a statistical analysis of these variations.
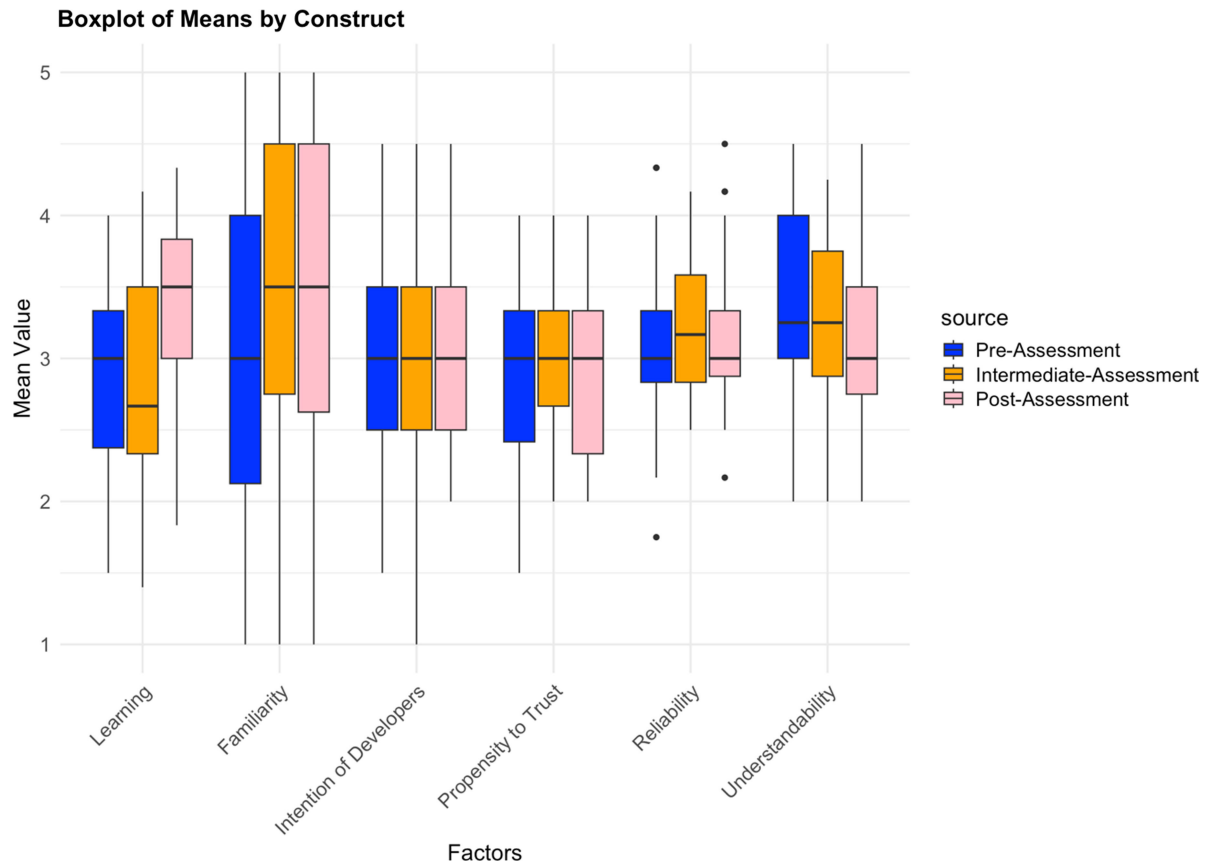
Figure 3. Boxplot of means by construct

Table 3. P-values and direction of variation per construct.

Construct	P-value	Variation direction
Learning	0.01 *	↑ (Increase)
Familiarity	0.10.	↑ (Increase)
Intention of Developers	1.00	(No Change)
Propensity to Trust	0.74	(No Change)
Reliability	0.59	(No Change)
Understandability	0.09.	↓ (Decrease)

5.3. Statistical analysis

Observing non-normal distribution through the Shapiro-Wilk test, the significance of the variations was tested through the Wilcoxon test (Janez, 2006). The results are shown in Table 3. The variance in Learning between the Pre-Assessment and the Post-Assessment was proved statistically significant ($p <$ 0.05), while *Familiarity* and *Understandability* variances were not statistically significant ($p > 0.05$) but still exhibited small p-values ($p = 0.109$ and $p = 0.094$ respectively). The framework demonstrated a statistically significant impact on students' self-assessment regarding their *Learning* about trust in AI. This demonstrates that the proposed framework is effective in the case of design students, when there is the goal of increasing awareness on trust-related topics. Conversely, no significant variation was observed following the design activity, revealing that design activities alone are not enough to impact trust learning.

Familiarity increased significantly following the design activity, likely as a result of the hands-on Design experience obtained through the use of ChatGPT4o-mini. This observation suggests that integrating generative AI-based design activities into classroom settings can enhance students' perceived knowledge of the system. In contrast, *Understandability* exhibited a gradual decline throughout the experiment, potentially reflecting a recalibration of students' initial assumptions as they gained a clearer view of the

system's capabilities and limitations. In other words, while greater *Familiarity* may have improved students' engagement with and perceived relevance of genAI in design, it also appears to have heightened their awareness of the complexities involved and their own gaps in understanding.

The fact that the proposed framework statistically increases *Trust Learning*, together with the achievement of a more critical view towards the use of genAI in design, makes us conclude that the proposed experiment can be a value-added activity to be brought in design classes. The results of the experiment and the proposed framework are a first step for the Design Community to increase the research and educational focus on trust. In the next section we conclude our contribution also highlighting some next steps of research in this direction.

6. Conclusions

This study has demonstrated the significant impact of the use of the framework on *Trust Learning*. By analyzing the students' self-assessed levels of knowledge on trust throughout the experiment, we have provided evidence supporting the hypothesis that only practical design activities do not alter the students perceived learning on trust. Thus, we suggest the importance of a mixed approach where hands-on design tasks accompany lectures and workshops. The impact of the lecture-workshop by itself is still to be assessed. However, notable variations in *Familiarity* and *Understandability* were observed during the design activities, suggesting that these changes are unlikely to be replicated solely through the lecture-workshop session. Future research could explore if the identified variations and the application of the framework affect the students' actual behavior towards LLMs. While this research offers valuable insights, it is important to acknowledge the lack of prior studies evaluating learning through self-assessment on learning outcomes. Thus, future research could explore this approach, analyzing its strengths and limitations, that could potentially influence the study's findings. It could also incorporate behavioral measures to accompany self-reports. Another limitation of the research is the absence of a control group, caused by the limited number of participants. For the same reason, it was not possible to conduct an analysis of the level of trust and trust learning based on the role taken by the students during the design activity (*Representative* or *Execution Team*). The paper does not analyze the relationship between trust and performance in design. We have preliminary data on this topic but aim to collect additional data to clarify whether such a relationship exists. Specifically, within the same theme, it will be interesting to investigate whether trust has differing effects on convergent and divergent phases, and whether these differences are associated with varying AI performance across distinct stages of the design process (Chiarello et al., 2024). Another promising avenue of exploration, particularly in the context of education, is to understand if and how the presented framework can influence the dynamics of distrust and overtrust.

With the increasing presence of the AI in all human activities, addressing how future generations will trust it will remain critical. In increasing our understanding of this topic, we aim to contribute to the discussion on how trust in AI can be shaped and supported through education and practical experience.

References

Campean, F., Yildirim, U., Korsunovs, A., & Doikin, A. (2024). Extending the function failure modes taxonomy for intelligent systems with embedded AI components. *Proceedings of the Design Society*, 4, 1949–1958. https://doi.org/10.1017/pds.2024.197

Chiarello, F., Barandoni, S., Škec, M. M., & Fantoni, G. (2024). Generative large language models in engineering design: opportunities and challenges. *Proceedings of the Design Society*, 4, 1959–1968. https://doi.org/10.1017/pds.2024.198

Clarke, R. E., Briggs, J., Armstrong, A., MacDonald, A., Vines, J., Flynn, E., & Salt, K. (2021). Socio-materiality of trust: co-design with a resource limited community organisation. *CoDesign*, 17(3), 258–277. https://doi.org/10.1080/15710882.2019.1631349

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, 14(2), 627–660. https://doi.org/10.5465/annals.2018.0057

Hart, S., Jan Hultink, E., Tzokas, N., & Commandeur, H. R. (2003). Industrial companies' evaluation criteria in new product development gates. *Journal of Product Innovation Management*, 20(1), 22–36. https://doi.org/10.1111/1540-5885.201003

Hu, X., Liu, A., & Dai, Y. (2024). Combining ChatGPT and knowledge graph for explainable machine learning-driven design: a case study. *Journal of Engineering Design*, 1–23. https://doi.org/10.1080/09544828.2024.2355758

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

Körber, M. (2019). Theoretical considerations and development of a questionnaire to measure trust in automation. In *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018) Volume VI: Transport Ergonomics and Human Factors (TEHF), Aerospace Human Factors and Ergonomics* 20 (pp. 13–30). Springer International Publishing.

Larson, L., & DeChurch, L. A. (2020). Leading teams in the digital age: Four perspectives on technology and what they mean for leading teams. *The leadership quarterly*, 31(1), 101377. https://doi.org/10.1016/j.leaqua.2019.101377

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human factors*, 46(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392

Liliana, L. (2016, November). A new model of Ishikawa diagram for quality assessment. In *Iop conference series: Materials science and engineering* (Vol. 161, No. 1, p. 012099). IOP Publishing. https://doi.org/10.1088/1757-899X/161/1/012099

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of management review*, 20(3), 709–734. https://doi.org/10.2307/258792

Mcknight, D. H., Carter, M., Thatcher, J. B., & Clay, P. F. (2011). Trust in a specific technology: An investigation of its components and measures. *ACM Transactions on management information systems (TMIS)*, 2(2), 1–25. https://doi.org/10.1145/1985347.1985353

Meron, Y., & Araci, Y. T. (2023). Artificial intelligence in design education: evaluating ChatGPT as a virtual colleague for post-graduate course development. *Design Science*, 9, e30. https://doi.org/10.1017/dsj.2023.28

Mollick, E., (2024). Co-Intelligence: Living and Working with AI. *Random House UK*.

Novelli, C., Taddeo, M., & Floridi, L. (2023). Accountability in artificial intelligence: what it is and how it works. *AI & SOCIETY*, 1–12. https://doi.org/10.1007/s00146-023-01635-y

Oxley, N. L., Dzindolet, M. T., & Paulus, P. B. (1996). The effects of facilitators on the performance of brainstorming groups. *Journal of social behavior and personality*, 11(4), 633–646.

Paschen, J., Kietzmann, J., & Kietzmann, T. C. (2019). Artificial intelligence (AI) and its implications for market knowledge in B2B marketing. *Journal of business & industrial marketing*, 34(7), 1410–1419. https://doi.org/10.1108/JBIM-10-2018-0295

Razin, Y. S., & Feigh, K. M. (2023). Converging measures and an emergent model: A meta-analysis of human-automation trust questionnaires. *arXiv preprint arXiv:2303.13799*. https://arxiv.org/pdf/2303.13799

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of management review*, 23(3), 393–404. https://doi.org/10.5465/amr.1998.926617

She, J., Neuhoff, J., & Yuan, Q. (2021). Shaping pedestrians' trust in autonomous vehicles: an effect of communication style, speed information, and adaptive strategy. *Journal of Mechanical Design*, 143(9), 091401. https://doi.org/10.1115/1.4049866

Song, B., Zhu, Q., & Luo, J. (2024). Human-AI collaboration by design. *Proceedings of the Design Society*, 4(DESIGN 2024), 2247–2256. Cambridge University Press. https://doi.org/10.1017/pds.2024.227

Torrance, E. P. (1966). Torrance tests of creative thinking. *Educational and psychological measurement*. https://doi.org/10.1037/t05532-000

Trauer, J., Schweigert-Recksiek, S., Schenk, T., Baudisch, T., Mörtl, M., & Zimmermann, M. (2022). A digital twin trust framework for industrial application. *Proceedings of the Design Society*, 2, 293–302. https://doi.org/10.1017/pds.2022.31

Troshani, I., Rao Hill, S., Sherman, C., & Arthur, D. (2021). Do we trust in AI? Role of anthropomorphism and intelligence. *Journal of Computer Information Systems*, 61(5), 481–491. https://doi.org/10.1080/08874417.2020.1788473

Urban, M., Děchtěrenko, F., Lukavský, J., Hrabalová, V., Svacha, F., Brom, C., & Urban, K. (2024). ChatGPT improves creative problem-solving performance in university students: An experimental study. *Computers & Education*, 215, 105031. https://doi.org/10.1016/j.compedu.2024.105031

Wijngaards, N. J., Boonstra, H. M., & Brazier, F. M. (2004). The role of trust in distributed design. *AI EDAM*, 18(2), 199–209. https://doi.org/10.1017/S0890060404040132