2024 IEEE Conference on Artificial Intelligence (CAI)

Encouraging Trust in AI-Powered Teaching Tools: Ranking Design Principles

Lee Peney
The Hague University of Applied Sciences
Zoetermeer, The Netherlands
L.L.L.Peney@student.hhs.nl

Raoul Dernee
The Hague University of Applied Sciences
Zoetermeer, The Netherlands
R.W.Dernee@student.hhs.nl

Hani Alers
The Hague University of Applied Sciences
Zoetermeer, The Netherlands
H.Al-Ers@hhs.nl

***Abstract*— The advent of Artificial Intelligence (AI) in education promises transformative changes, but its effective adoption relies on the establishment of trust. Drawing from various scientific articles, this study identified seven design principles that influence students' trust towards AI-powered teaching tools. These principles are "Privacy", "Intelligence," "Fairness", "Controllability", "Engagement", "Transparency", and "Friendliness". A quantitative survey, involving students from the Hague University of Applied Sciences, was employed to rank these principles based on their perceived importance. The results revealed students find all principles important except for "Friendliness". Gender-based analysis indicated females' pronounced emphasis on "Fairness", "Friendliness" and "Engagement" compared to males. Further analysis revealed students in ICT related fields value "Privacy" significantly more than their non-ICT related counterparts and "Fairness" significantly less.**

Keywords— trust, encourage trust, AI, education, design principle, chatbot.

I. Introduction

In the realm of education, the integration of Artificial Intelligence (AI) has emerged as a substantial change, reshaping traditional educational methods and systems. The potential of AI in revolutionizing education is undeniable, with its ability to offer personalized and efficient learning experiences [1]. However, this transformation is not without its challenges.

Central to this change in AI in education is the concept of trust. Trust plays a foundational role in traditional educational settings, particularly in the dynamics of teacher student relationships [2]. This significance of trust extends to the domain of AI, where the abrupt introduction of AI tools, such as the ChatGPT application, has elicited concerns and demands from educators [3]. The perceptions of university teaching tools towards such applications underscore the importance of trust and ethical considerations in the successful adoption of AI technologies in educational settings. Trust in AI, especially in educational settings, is of significant importance for effective learning outcomes and the successful adoption of AI-based educational tools [4].

Resistance to trusting new technologies is not a new phenomenon. One historical precedent is the adoption of surgical instruments in the 19th century [5]. These tools were also initially met with skepticism, which significantly slowed down their widespread acceptance. The correlation between the past adoption of surgical instruments and the current implementation of AI tools emphasizes enduring challenges and principles tied to the introduction of novel technologies. Trust, safety, education, transparency, regulation, and continuous refinement emerge as recurring themes in historical and modern technological progress, underscoring the significance of addressing these elements for a successful integration into society.

The ethical implications of AI technologies, especially in the context of education, cannot be overlooked. The surge in the use of online exam supervision technologies during the COVID-19 pandemic brought forth a plethora of ethical concerns, ranging from academic integrity to privacy and autonomy [6]. Such concerns reappear in the realm of AI technologies, where the potential of AI to produce ethical arguments raises questions about the integrity of academic work [7].

Furthermore, the adoption of AI-based chatbots in higher education institutions is influenced by a range of factors, including trust. Research indicates that the effectiveness of chatbot adoption hinges on addressing trust and ethical considerations [8]. This sentiment is reinforced in the proposal by the international conference on higher education advances. They propose an ethically aligned curriculum for computer sciences and information technology specialties, emphasizing the integration of ethical considerations in the face of global digitalization [9].

In synthesizing the insights from these articles, a pattern emerges. While some articles emphasize the importance of trust in both traditional and AI-integrated educational settings [1][4], others highlight the ethical challenges and implications of AI technologies [5][7]. The span of publication dates from the 19th century to 2023, further underscores the recurring importance of trust and ethics in the context of technological advancements in education.

While AI promises vast changes in educational methodologies, the factor of trust remains a big issue. For AI-powered teaching tools to be embraced by students, it is imperative to address this trust issue. Various scientific articles and reports have stated principles aimed at enhancing trust in AI systems. This article seeks to understand their importance to students. Furthermore, we aim to prioritize these principles, providing a guide for future design and development decisions for AI-Tools used for teaching.

Based on the information gathered the following question was established. How can we make AI-powered teaching tools more trustworthy to students based on design principles, and how would students rank the most effective design principles for achieving this trust?

To answer this question two sub questions were established, those being. What general design principles exist that increase user trust in online tools able to collect private

DOI 10.1109/CAI59869.2024.00096

data about them? And which of these design principles do users find more important? The first sub question was answered with reviewing relevant literature. The second sub question was answered with the use of a questionnaire.

II. Design principles

Several design principles for encouraging trust in AI-powered educational tools from relevant literature are presented below. These principles are categorized into various subsections, each highlighting a unique aspect imperative for enhancing user experience and trustworthiness.

These design principles supported by reputable sources including the European Commission, U.S. Department of Education, and notable publications and corporations like IBM and Google, lay a solid foundation for trust in AI-powered educational tools.

A. Privacy

The importance of upholding user privacy cannot be overstated. In an era where data is a valuable commodity, safeguarding it is paramount. The focus here lies not only on maintaining data quality but also on rigorously regulating access to user data, ensuring that only authorized personnel have the privilege to interact with this sensitive information. This commitment to privacy is well-documented in the literature [10, 11, 13, 14], underlining the adherence to established best practices and ethical guidelines.

B. Fairness

In the quest to create an inclusive educational environment, the principles of fairness, accessibility, and participation stand as pillars of strength. Every user, irrespective of their background or capabilities, should feel welcomed and accommodated. This commitment to inclusivity does not just foster a sense of belonging but also contributes to equity among all users [10, 13, 14].

C. Controllability

Granting users control over their personal information is not just a gesture; it is an essential aspect of ethical AI education. Empowering users to view and modify their data as they see fit reflects the respect for their autonomy and a commitment to transparency. These actions resonate with the principles outlined in relevant literature [12, 13], which highlight the user-centric approach to data management.

D. Transparency

Unlike conventional web browsing or app usage, interactions with AI can often be sporadic and fragmented. This quirk necessitates an additional layer of transparency. AI-powered educators must provide clear indicators of the user's current position within the educational journey, eliminating any confusion and fostering trust. The necessity for this transparency is eloquently underscored in literature [11, 12, 13], emphasizing the importance of making the user's experience as seamless and comprehensible as possible.

E. Intelligence

For AI educators to be considered trustworthy, they should embody a sense of intelligence that surpasses human capabilities. This means not only processing data swiftly but also doing so with a high degree of accuracy. Employing visualizations can assist in conveying this intelligence, making users feel that they are interacting with a system that is not only capable but also responsive [12, 13].

F. Friendliness

Interactions with AI educators should transcend mere functionality. They should evoke a sense of friendliness, aiming to forge a genuine relationship with the user, which can make interactions feel more personal and less mechanical. The adoption of this approach aligns with the idea of building rapport [11, 12].

G. Engagement:

Beyond just imparting knowledge, AI educators should also inspire, encourage, and recognize user achievements. This sense of encouragement is vital for engagement and motivation. AI educators should be designed to promote a collaborative environment where users have the freedom to suggest ideas and make decisions, thereby enhancing their engagement and trust in the educational process [11, 12, 14].

III. Methodology

The methodology outlined here has been structured to systematically prioritize design principles that encourage trust among students towards AI-powered teaching tools.

A comprehensive questionnaire was constructed. This questionnaire is meant to collect two types of data. Firstly, it collects demographic data, capturing insights related to the participants' age, gender, and education. Secondly, the questionnaire prompts students to rank the design principles according to their perceived importance. Each principle is ranked on 1 – 7 Likert scale with 1 being not important and 7 being very important.

To ensure a diverse sample, participants were randomly selected from The Hague University of Applied Sciences. This strategy allowed for a mix of ages (18-29), genders, and educational backgrounds. Participants were recruited randomly picking a person, walking up to said person and asking them to fill out the questionnaire. The questionnaire was filled in by thirty-nine participants.

IV. Results

A. Differences in Importance Across Design Principles

A repeated measures ANOVA was conducted to evaluate the variance in importance ratings across the seven design principles. The analysis revealed a statistically significant difference in the ratings across the different principles, $F(6, 48) = 8.16$, $p < .001$.

To identify which principles were rated significantly different from each other, a post-hoc test with Holm's correction was performed. The test revealed several pairs of principles that were significantly different in their ratings. Notably, “Friendliness” exhibited significant differences when compared to most other principles:

“Privacy” vs. “Friendliness” ($p < .001$, $d = 1.319$)

“Fairness” vs. “Friendliness” ($p < .001$, $d = 1.171$)

“Controllability” vs. “Friendliness” ($p = .001$, $d = 0.836$)

“Transparency” vs. “Friendliness” ($p < .001$, $d = 0.911$)

“Intelligence” vs. “Friendliness” ($p < .001$, $d = 1.338$)

“Friendliness” vs. “Engagement” ($p = .001$, $d = -0.855$)

No other significant differences were found,

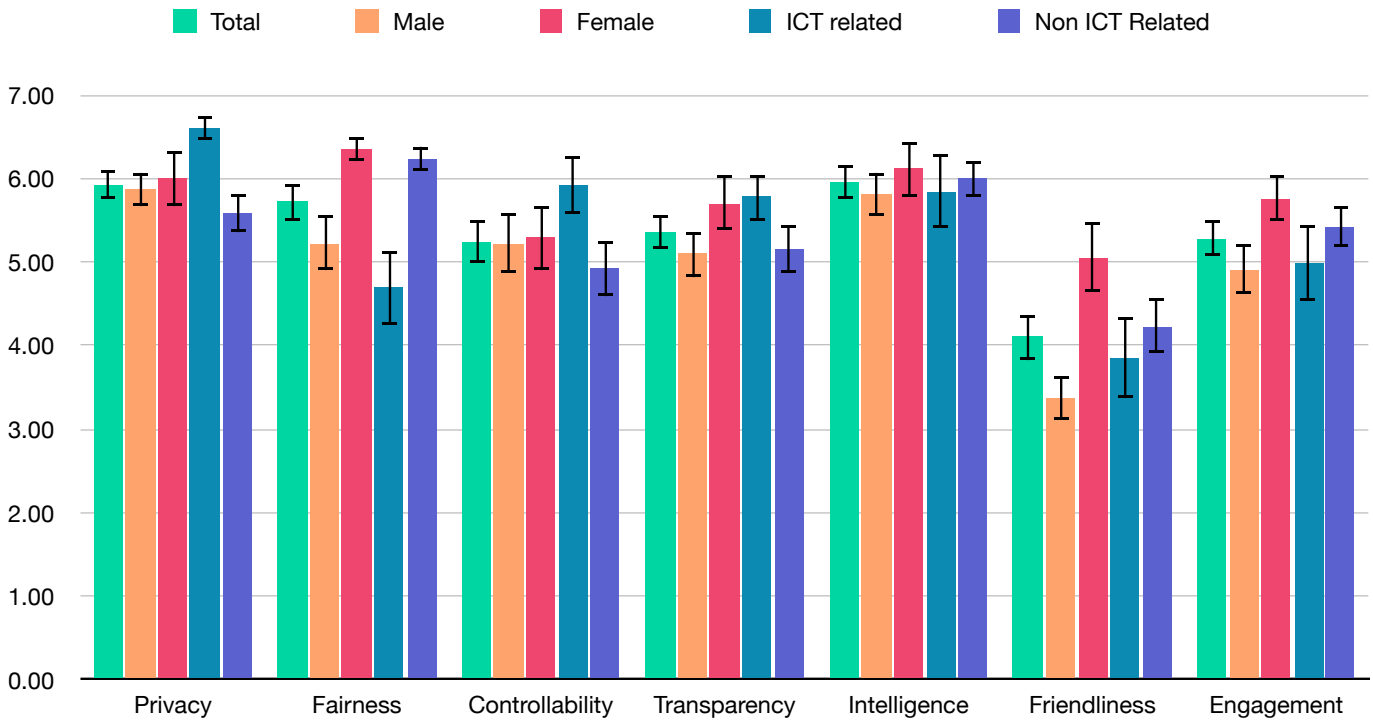

Figure 1. Bar chart of trust scores for the design principals across different demographics. Error bars represent the standard error of the mean.

The significant p-values and large effect sizes (Cohen's d) indicate that "Friendliness" is consistently rated as less important compared to other design principles. These findings suggest that the perceived importance of design principles is not uniform. Since no other significant differences were found all other principles are of equal importance to students

During the data analysis phase, it was determined there were significant differences in the rankings between different demographics. Gender and field of study play a big role in the ranking of the design principles. The age of the participants, ranging from 18 to 29 years old, was determined not to have an effect on the ranking.

B. *Effect of Gender on Evaluation of Design Principles*

A series of independent sample t-tests were conducted to compare the ratings of different design principles between male and female students. There was a significant difference in the ratings for "Fairness" ($t(48) = -2.95$, $p = .006$), "Friendliness" ($t(48) = -3.61$, $p = .001$), and "Engagement" ($t(48) = -2.03$, $p = .049$) between the two genders.

To gauge the practical significance of these findings, Cohen's d was calculated for each significantly different principle. The effect sizes for "Fairness" ($d = -0.952$), "Friendliness" ($d = -1.167$), and "Engagement" ($d = -0.656$) suggest medium to large effects.

The negative sign of the d values indicates that female students rated these principles higher than male students. Specifically, the large effect sizes for "Fairness" and "Friendliness" suggest that the observed differences are not only statistically significant but also practically relevant. This finding implies that gender plays a substantial role in how students evaluate these particular design principles, warranting further investigation into the underlying factors.

C. *Effect of ICT-Related Study Programs on Evaluation of Design Principles*

A series of independent sample t-tests were conducted to compare the ratings of different design principles between students in ICT-related and non-ICT-related study programs. There was a significant difference in the ratings for "Privacy" ($t(48) = 3.04$, $p = .004$) and "Fairness" ($t(48) = -4.18$, $p < .001$) between the two groups.

To assess the practical significance of these findings, Cohen's d was calculated for each significantly different principle. The effect sizes for "Privacy" ($d = 1.033$) and "Fairness" ($d = -1.421$) suggest large effects.

The direction of the d values indicates that students in ICT-related programs rated "Privacy" higher, while students in non-ICT-related programs rated "Fairness" higher. The large effect sizes for both principles indicate that these differences are not only statistically significant but also practically relevant. This suggests that the field of study significantly influences how students evaluate these specific design principles. Further research is needed to explore the factors contributing to these differences. These results are visualized in Figure 1.

V. Conclusion

In conclusion, the findings of this study shed light on the preferences and priorities of students in ranking design principles for AI teaching tools. The questionnaire, involving thirty-nine participants, provided valuable insights into the considerations that students take into account when evaluating AI teaching tools. The investigation reveals distinct patterns in how distinct groups perceive the value of various design principles in AI teaching tools.

Moreover, the analysis pointed to a hierarchy of design principles, such as "Friendliness", consistently being rated significantly lower relative to others. This suggests that students do indeed have significant differences in their preferences, and certain attributes may be deemed less critical to trustworthy AI teaching tools.

Gender emerged as a significant factor, with significant differences in how male and female students prioritize aspects of AI tool design. Females tended to hold principles like "Fairness", "Inclusivity", and the friendly nature of the tools in higher regard, suggesting that gender-related perspectives and expectations may influence the perception of AI teaching tools.

When assessing the impact of academic background, students in ICT fields showed a significant lean towards valuing "Privacy" within AI-powered teaching tools, possibly stemming from a more profound awareness of the complexities involved in data security. Conversely, those from non-ICT backgrounds placed greater importance on "Fairness", possibly reflecting a broader concern for equitable and accessible educational environments.

These patterns underscore the complexity of designing AI teaching tools that are universally accepted. Acknowledging the substantial influence of both gender and educational background on students' preferences can guide developers and educators in creating more tailored and effective AI teaching environments. The evidence advocates for continued exploration into these preferential divides to foster AI tools that resonate with the intricate and varied fabric of the student body.

VI. Discussion

The research in Artificial Intelligence (AI) for education has far-reaching effects on society and the scientific community. It significantly contributes to our understanding of building trust through design principles, setting the stage for further advancements in human-computer interaction and AI ethics research. This approach not only draws attention from academic disciplines such as computer science, education, psychology, and ethics but also fosters collaboration across these fields.

Moreover, the findings have the potential to influence policy development in the education sector, especially concerning AI usage, data privacy, and inclusivity. By emphasizing ethical considerations in AI development, the research contributes to ongoing discussions within the scientific community about responsible AI. This, in turn, encourages international collaboration and information exchange on AI in education, benefiting countries interested in implementing similar technologies.

Overall, this research has the potential to create a positive impact on both society and the scientific community by advancing knowledge, improving education, and promoting ethical and responsible AI development. It can play a pivotal role in shaping the future of AI in education, making it more inclusive, transparent, and trustworthy for students and educators. However, it is important to acknowledge certain limitations in the study. The relatively small sample size of thirty-nine participants and limited to one university, which may not capture the full diversity of student opinions. In future research, expanding the participant pool to include a more significant and diverse group could yield more comprehensive and generalizable insights.

Furthermore, the research focused on students, neglecting the perspectives of teachers on AI teaching tools. Future studies should aim to determine which design principles make an AI-powered teacher more trustworthy among educators, providing a more holistic view of the topic. The questionnaire used in this research, while valuable for understanding student perceptions, can be adapted for teacher-focused investigations to further enrich the body of knowledge in this area. By encompassing multiple institutions and various demographic backgrounds, future studies can provide a more comprehensive and nuanced understanding of the role of AI in education from both student and teacher perspectives.

VII. References

[1] Harry, A. (2023). Role of AI in Education. *Interdiciplinary Journal and Hummanity (INJURITY)*, *2*(3), 260-268.

[2] Jeong, S., Irby, B. J., Pugliese, E., Boswell, J., & Kappler Hewitt, K. (2018). Editor's overview: exploring mentoring relationships of various populations. *Mentoring & Tutoring: Partnership in Learning*, *26*(5), 477-481.

[3] Alarcón Llontop, L. R., Pasapera Ramírez, S., & Torres-Mirez, K. (2023). The ChatGPT Application: Initial Perceptions of University Teachers. *LACCEI Journal Test*, *2*(6).

[4] Bilad, M. R., Yaqin, L. N., & Zubaidah, S. (2023). Recent progress in the use of artificial intelligence tools in education. *Jurnal Penelitian dan Pengkajian Ilmu Pendidikan: e-Saintika*, *7*(3), 279-314.

[5] Swayne, J. G. (1869). On the Use of Obstetric Instruments. *British medical journal*, *1*(439), 487.

[6] Coghlan, S., Miller, T., & Paterson, J. (2021). Good proctor or "big brother"? Ethics of online exam supervision technologies. *Philosophy & Technology*, *34*(4), 1581-1606.

[7] Zohny, H., McMillan, J., & King, M. (2023). Ethics of generative AI. *Journal of medical ethics*, *49*(2), 79-80.

[8] Mohd Rahim, N. I., A. Iahad, N., Yusof, A. F., & A. Al-Sharafi, M. (2022). AI-based chatbots adoption model for higher-education institutions: A hybrid PLS-SEM-neural network modelling approach. *Sustainability*, *14*(19), 12726.

[9] Libin, A. (2020, May). Integrated disciplines and future competencies: A blueprint for ethically aligned curriculum for IT, CS, ITC & beyond. In *6th International Conference on Higher Education Advances (HEAd'20)* (No. 30-05-2020, pp. 1245-1251). Editorial Universitat Politècnica de València.

[10] European Commission, Directorate-General for Education, Youth, Sport, and Culture. (2022). Ethical guidelines on the use of artificial intelligence (AI) and data in teaching and learning for educators.

[11] U.S. Department of Education, Office of Educational Technology, Artificial Intelligence and Future of Teaching and Learning: Insights and Recommendations, Washington, DC, 2023.

[12] Guo, Y., Wang, J., Wu, R., Li, Z., & Sun, L. (2022). Designing for trust: A set of design principles to increase trust in chatbot. *CCF Transactions on Pervasive Computing and Interaction*, *4*(4), 474-481.

[13] IBM. (2022). Building trust in AI. Retrieved from ibm.com: https://www.ibm.com/impact/ai-ethics

[14] Google. (2023). Responsibility: Our principles. Retrieved from ai.google.com: https://ai.google/responsibility/principles/