GOOD SYSTEMS
A UT Grand Challenge

Student-centered AI for intelligent tutoring systems:

Empowering students with agency, metacognition, self-efficacy, trust, and usability

Eunhye Grace Ko[1], Dr. Joan E Hughes[1], Dr. Ken Fleischmann[2]
[1]Learning Technologies, College of Education, [2]School of Information

TEXAS Education
The University of Texas at Austin
College of Education

The University of Texas at Austin
School of Information

Introduction

AI-based intelligent tutoring systems (ITS) are being integrated into education at an accelerating pace, but ethical challenges of privacy, bias, and discrimination arise. To address these challenges, empirical research applying human-centered AI design techniques such as explainable AI (XAI) and human-in-the-loop features is critical. XAI promotes transparency and trust, while the human-in-the-loop feature increases user control and personalization. However, empirical studies of these features in education are rare. Through an interdisciplinary approach to learning science, HAI, and AI ethics, this study aims to design and measure the impact of a human-centered AI intelligent tutoring system (HAI-ITS) prototype. In community college developmental math course context, the study intends to provide instructional design implications for planning ITSs with human-centered AI design elements while addressing trust and agency issues of AI integration and self-efficacy and metacognition.

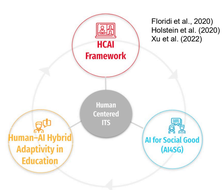
Research Question

Based on a value-sensitive design approach, what is the design of ITS that incorporates AI ethics, human-centered AI, and human–AI hybrid adaptivity?

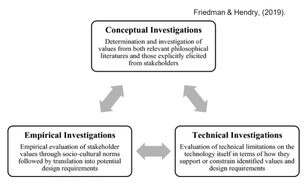
1. What are the core human values that are addressed in the HAI-ITS prototype design process?
2. What results from the conceptual, empirical, and technical investigations?.
3. How can the investigation results be reflected in the prototype design?

Methods

Conceptual Framework



Methodological Framework



Conceptual Investigation

- Analysis of AI ethics, human-centered AI design theories, and AI-student interaction scenarios

Technical Investigation

- Analysis of human-AI interaction guidelines from industry
- Interviews with a developer and data scientist

Empirical Investigation

- Focus group/ individual interviews, Zoom 30m-1h
- Direct Stakeholders: 7 Community College Students, 5 Instructors
- Indirect stakeholders: 1 AI ethics researcher, 1 UX designer, 1 instructional designer

Results

Empirical investigation

- Identified student and instructor pain points and user needs in relation to explainable AI and human-in-the-loop features while interacting with AI-based intelligent tutoring systems
- **Pain points:** 14 instructor and student common, 5 student-specific, and 4 instructor-specific. Examples:
  Common: assessment interpretation/ clarity
  Student: lacking concept connection
  Instructor: trust building for AI, communicate learning science
- **User needs:** 7 Human-in-the-loop, 4 Explainable AI related needs identified. Examples:
  XAI1: setting up student expectations for AI's functionalities
  XAI2: capacity to make connections between concepts
  Human-in-the-loop 1: capacity to confirm understanding
  Human-in-the-loop 2: communicating struggle situations
- Revealed 9 **Value Tensions** impacting the design of prototype. Example: Transparency vs Trust - Mixed feelings among students and instructors about disclosing the confidence level of AI-based learning system. Some see value in sharing AI's imperfections as it can improve trust, while others feel that educational software should be accurate before it's used.

Conceptual investigation

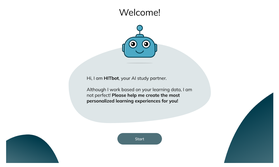
- Determined direct and indirect human values that can impact the design of human-centered AI intelligent tutoring system (HAI-ITS) prototype (Friedman et al.,2006; Cheng & Fleischmann, 2010). Direct value examples: Ownership, Privacy, Universal Usability, Trust, Autonomy, Accomplishment
- Investigated various AI ethics frameworks and selected AI for Social Good (AI4SG; Floridi et al., 2020) framework as conceptual framework for its practical applicability.
- Analyzed AI-student interaction scenarios (Seo, 2021; Holstein et al., 2019; Conani et al., 2019) and identified 20 scenarios that can be reference to the design process

Technical investigation

- By analyzing human-centered AI design provided by Apple, Google, and Microsoft. classified 9 human-centered AI design techniques that can be applied to the design of student-centered ITS. Example: Add explanations in readily comprehensible language in context of personalized learning (Google, Apple). Reveal context as to where explanations are provided, explain why question is given at the moment (MS).
- By conducting interviews with a developer and a data scientist, ensured the prototype's technical feasibility.

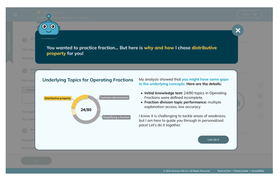
Integration of findings from Conceptual, Technical and Empirical investigation to prototype design

Prototype examples



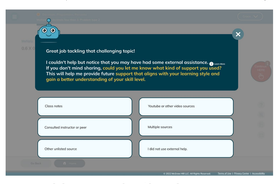
Introduction

XAI - Helping students understand how AI system works to benefit learning

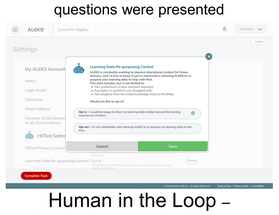


XAI - Showing why & how the questions were presented

Human in the Loop -
By communicating students' confidence over mastery, AI can provide better personalized learning path



Human in the Loop -
Instructors think external help is cheating and harms AI accuracy. Encouraging students to disclose external help can prevent cheating, improve AI's accuracy, and provide personalized learning support.



Human in the Loop –
Bring to data-ownership issues in edtech to students' attention and allow students to have better control over their privacy.

Further Studies

As a follow-up study, the prototype is currently being implemented and tested with community college students taking developmental math courses. The study aims to answer research question on "How do HAI features (human-in-loop & explainable AI) in the ITS prototype impact students' perceptions of their learner agency, metacognition, self-efficacy, usability of ITS, and trust in ITS?"

The preliminary findings from 70 participants showed:

- Human-in-the-loop feature to receive additional help while learning was a significant predictor of Learner Agency ($p < .01$), Metacognition ($p < .01$), Self-efficacy ($p < .01$), Usability ($p < .01$), and Trust ($p < .01$).
- Explainable AI feature was a significant predictor of Usability ($p < .01$).
- Human-in-the-loop feature to control privacy was a significant predictor of Usability ($p < .01$).
- In terms of individual differences, technology self-efficacy was a significant predictor of Trust ($p < .05$), Metacognition ($p < .05$).

The study is continuously collecting data in Spring 2023 and the final result will be updated accordingly.

Acknowledgments

The authors gratefully acknowledge the support of the Good System for providing mentorship and fellowship that facilitated our research. We would also like to thank the community college students and instructors who provided valuable insights into the topic. Finally, we extend our appreciation to the diverse experts who generously shared their professional insights with us, helping to enrich the scope and depth of this study.

References

- Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019, May 1). *Guidelines for Human-AI Interaction*. CHI 2019. https://www.microsoft.com/en-us/research/publication/guidelines-for-human-ai-interaction/
- Apple. (n.d.). Human interface guidelines for machine learning. https://developer.apple.com/design/ human-interface-guidelines/machine-learning/ overview/introduction/
- Cheng, A. S., & Fleischmann, K. R. (2010). Developing a meta-inventory of human values. Proceedings of the *American Society for Information Science and Technology, 47*(1), 1-10.
- Conati, C., Barral, O., Putnam, V., & Rieger, L. (2021). Toward personalized XAI: A case study in intelligent tutoring systems. *Artificial Intelligence*. https://doi.org/10.1016/J.ARTINT.2021.103503
- Floridi, L., Cowls, J., King, T. C., & Taddeo, M. (2020). How to design AI for social good: seven essential factors. *Science and Engineering Ethics, 26*(3), 1771-1796. https://doi.org/10.1007/s11948-020-00213-5.
- Friedman, B., & Hendry, D. G. (2019). Value sensitive design: Shaping technology with moral imagination. Mit Press.
- Google. (n.d.). *Google PAIR. People + AI Guidebook: Explainability + Trust* https://pair.withgoogle.com/chapter/explainability-trust/
- Holstein, K., McLaren, B.M., Aleven, V. (2019). Designing for Complementarity: Teacher and Student Needs for Orchestration Support in AI-Enhanced Classrooms. In: Isotani, S., Millán, E., Ogan, A., Hastings, P., McLaren, B., Luckin, R. (eds) *Artificial Intelligence in Education (AIED) 2019* (11625). Springer, Cham. https://doi.org/10.1007/978-3-030-23204-7_14
- Holstein, K., Aleven, V., Rummel, N. (2020). A Conceptual Framework for Human–AI Hybrid Adaptivity in Education. In: Bittencourt, I., Cukurova, M., Muldner, K., Luckin, R., Millán, E. (eds) *Artificial Intelligence in Education. AIED 2020*(12163). Springer, Cham. https://doi.org/10.1007/978-3-030-52237-7
- Seo, K., Tang, J., Roll, I., Fels, S., & Yoon, D. (2021). The impact of artificial intelligence on learner–instructor interaction in online learning. *International Journal of Educational Technology in Higher Education, 18*(1), 1-23. https://doi.org/10.1186/s41239-021-00292-9
- Xu, W., Dainoff, M. J., Ge, L., & Gao, Z. (2022). Transitioning to Human Interaction with AI Systems: New Challenges and Opportunities for HCI Professionals to Enable Human-Centered AI, *International Journal of Human–Computer Interaction*, DOI: 10.1080/10447318.2022.2041900

The University of Texas at Austin
Research Development
Vice President for Research, Scholarship and Creative Endeavors