

Impact of AI Explainability on College Students' Trust in AI

Swathi Sridhar, Katelyn Batoy,
Taylor Silva-Kaiser, Tram Le, David Schuster
San José State University

Introduction

- College students and educators have begun to utilize AI applications in many learning environments, such as classrooms and remote learning.
- Fostering appropriate trust has become increasingly important due to the convenience that artificial intelligence (AI) and large language models (LLMs) provide.

Background

- With this AI expansion, trust emerges as the cornerstone of AI usage, traversing different fields.
- AI has finite capabilities—i.e., its abilities do not always align with the users' expectations.
- Trust Calibration: the balance of excess trust and distrust in AI (Turner et al., 2020).
 - This term suggests that humans adjust their prior expectations and schemas of AI to correspond with AI's reliability.

Research Problem

- Trust often plummets to extremes where people either over-rely on AI or primarily focus on manual processes, neglecting AI's full potential.

SJSU SAN JOSÉ STATE
UNIVERSITY

THEORY

This presentation aims to provide a roadmap for researchers who want to investigate AI's impacts on college education, by focusing specifically on the level of AI explainability.

vectrlab.net

Research Purpose

- This methodological approach will help investigators discover foundational insights regarding appropriate AI trust in academic settings and encourage future experimental research in this domain.

Hypotheses

1. Higher levels of trust in AI will positively correlate with high levels of AI explanation,
2. A main effect will exist between critical thinking and trust in AI/automation, and
3. An interaction effect will exist between frequency of AI usage and trust on the levels of explanation provided by AI.

Broader Implications

- College education is a microcosm of the professional realm, as it parallels the diverse professional fields of study in the real world.
- Researchers must examine student trust and academic performance, as the findings from this proposed research could potentially inform how students eventually utilize AI in other environments, such as the workplace.

References

- Grimmelikhuijsen, S. (2023). Explaining why the computer says no: Algorithmic transparency affects the perceived trustworthiness of automated decision-making. *Public Administration Review*, 83(2), 241-262.
- Hoff, K. A., & Bashir, M. (2015). Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 57(3), 407-434. <https://doi.org/10.1177/0018720814547470>
- 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.46.1.50.30397>
- Osofsky, S., Schuster, D., Phillips, E., & Jentsch, F. (2013). Building Appropriate Trust in Human-Robot Teams. *Trust and Autonomous Systems: Papers from the 2013 AAAI Spring Symposium*.
- Roemmich, K., Schaub, F., & Andalibi, N. (2023). Emotion AI at Work: Implications for Workplace Surveillance, Emotional Labor, and Emotional Privacy. *Association for Computing Machinery*, 1-20. <https://dl.acm.org/doi/pdf/10.1145/3544448.3580950>
- Turner, A., Kaushik, M., Huang, M.-T., & Varanasi, S. (2020, May 18). Calibrating Trust in AI-Assisted Decision Making. *UC Berkeley School of Information*. <https://www.ischool.berkeley.edu/programs/mids/capstone>

 vectrlab @ sjsu

Poster layout adapted from Morrison (2019).