Felix Gröner

fgroener@asu.edu

Education

Arizona State University, USA *Expected Fall 2026*
Ph.D. Human Systems Engineering
Status: Ph.D. candidate, GPA 4.0
Committee: Erin K. Chiou (Advisor), D. Vaughn Becker, Nancy J. Cooke
Dissertation topic: Engineering AI agents for and validating LLM-driven simulations of synthetic participants for Human Factors research

Ulm University, Germany June 2024
M.Sc. Cognitive Systems
Thesis: The role of explainable artificial intelligence in trust and adherence to AI recommendations

Kiel University, Germany December 2022
M.A. International Politics and International Law
Thesis: Necessity of adapting International Humanitarian Law to cyber warfare

Kiel University, Germany December 2020
B.A. Computer Science and Political Science
Thesis: NATO's strategy to deter cyber attacks

Kiel University, Germany December 2019
B.Sc. Business Informatics
Thesis: Client's influence on the choice of project management technique in software development

Publications

Peer-reviewed

Gröner, F., Verma, A., & Bronowitz, J. (in press). Automatically Generating Interactive Learning Experiences with an LLM Driven Agentic Pipeline. *Second Annual Convening of the Learning Engineering Research Network.*

Gröner, F., & Chiou, E. K. (2024). Investigating the impact of user interface designs on expectations about large language models' capabilities. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting.* https://doi.org/10.1177/10711813241260399

Kraus, J. M., Merger, J., **Gröner, F.**, & Pätz, J. (2023). 'Sorry' says the robot: The tendency to anthropomorphize and technology affinity affect trust in repair strategies after error. *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction (pp. 436-441).* https://doi.org/10.1145/3568294.3580122

In Preparation

Gröner, F., Mancenido, M. V., Kim, N., Summers, E., & Chiou, E. K. (in preparation). *Increasing transparency of LLM systems does not always improve people's verification behavior and performance: Results from an empirical study of AI-assisted intelligence analysis reporting.*

Gröner, F., & Chiou, E.K. (under review). *Caveats of replacing human participants with LLM-powered simulated users: The case of a writing style preference inference simulation*.

Gröner, F., & Becker, D.V. (in preparation). *Subjective and objective metrics to explain perceptions of original and AI-generated abstract artworks*.

Gröner, F., Mancenido, M. V., & Chiou, E. K. (in preparation). Considering expertise as a factor in the design of synthetic test environments for human-AI research. In E. K. Chiou, D. S. Lange, J. H. Wong, & J. Marble (Eds.), *Advancements in Human-Agent Teaming Research Infrastructure: Testbeds, metrics, and concepts.* CRC Press.

Presentations

Gröner, F. & Chiou, E. K. (2026). Engineering approaches to improve human-LLM interaction: A scoping literature review. *AI Impact Pre-Summit @ ASU,* 29 January, Mesa, AZ.

Chiou, E. K., **Gröner, F.**, & Mancenido, M. V. (2025). Increasing transparency of LLM systems does not always improve people's verification behavior and performance: Results from an empirical study of an AI-assisted intelligence reporting task. Abstract submitted to the *69th HFES International Annual Meeting: ASPIRE*, 13-17 October, Chicago, IL.

Gröner, F., Mancenido, M. V., & Chiou, E. K. (2024). Multi-source AI Scorecard Table (MAST) for evaluating generative AI in worker-automation team tasks. Project update presented at *Center for Accelerating Operational Efficiency (CAOE) Annual Meeting,* 8-9 April, Washington, DC.

Professional Experience

The Polytechnic School, Arizona State University *2026*
Graduate Teaching Assistant
Responsibilities: Developing class material for a full-stack development capstone project

ADAPT Lab, Arizona State University *2025*
Graduate Research Assistant
Responsibilities: Conceptualizing and conducting pilot studies and literature reviews
Projects:

Semantic leakage is an issue when experimenting with synthetic datasets
- When generating synthetic datasets, prompting context can have undesired side-effects which threaten the validity of findings.
- I used mechanistic interpretability methods to identify the relevant attention heads. Ablating them did not reduce the issue.

People and LLMs learning each other's competencies
- Conducted a pilot study (n = 9) testing how well people and LLMs can assess their own and each other's likelihood of knowing the answer to the next question, given their performance history. Sensitivity to prompt design heavily influenced the initial findings.

Center for Accelerating Operational Efficiency *2023-2025*
Graduate Research Assistant
Responsibilities: writing grant application and research proposals, authoring papers
Projects:

Graphical user interfaces for directing expectations or LLMs (GUIDE-LLM)
- $132,000 awarded
- Developed a custom experimentation platform for human-LLM interaction research. This full-stack web application allows researchers to control the UI/UX of and record behavioral metrics in experiments of human-LLM interaction.
- Composed a set of tasks following the MAST framework to catalog misconceptions about LLM capabilities.

Multisource AI scorecard table (MAST) for evaluating generative AI in worker-automation team tasks
- $210,000 awarded
- Constructed an experimental setup in which participants (n = 304) solved an intelligence analysis task.
- Analyzed the data from different conditions that varied task difficulty, transparency, and stakes.

The Polytechnic School, Arizona State University *2023*
Graduate Teaching Assistant
Responsibilities: Teaching classes and recording videos for Introduction to Programming, Python, Raspberry Pi

Human Factors research group, Ulm University *2022-2023*
Undergraduate Research Assistant
Responsibilities: Designing and conducting research studies, project management
Projects:

The role of explainable artificial intelligence (XAI) in trust and adherence to AI recommendations

- Showing people (n = 404) either a global (linear surrogate) or local (LIME) explanation for an AI system's recommendation did not have as large of an effect as the participants' prior self-perceived expertise.