

Master's Thesis: Trust-Adaptive Intelligent Explanations in LLM-based Data Assistants

Supervisor: Ana-Maria Sîrbu (anamaria.sirbu@uni-passau.de)

Start date: as soon as possible

Motivation and Goals

Despite the remarkable capabilities of large language models (LLMs) such as GPT-5, Gemini, or Claude, their black-box nature often raises concerns about trustworthiness. Explainable AI (XAI) research shows that explanations can improve transparency and trust. Early studies on explanations (e.g., Gregor & Benbasat, 1999) introduced various explanation provision strategies, such as automatic (always displayed), user-invoked (on-demand), and intelligent (displayed when considered necessary by the system). Among these, intelligent explanations are particularly promising because they may reduce information overload while offering necessary insights without additional user effort. Recent work on trust-adaptive AI interventions (Srinivasan & Thomason, 2026) offers a promising basis for deciding *when* explanations should be shown *based on the user's current trust state*. For example, LLM-based data assistants could use users' reported trust as a contextual signal and provide explanations when trust is particularly low or high to address potential under- or overreliance.

The goal of this thesis is to implement and experimentally evaluate a trust-adaptive explanation mechanism that provides explanations based on users' dynamically reported trust. A basic LLM-Based Data Assistant prototype and its code can be provided as a starting point. The student should select and implement a brief measure of current trust, define transparent thresholds and decision rules for explanation provision, and integrate the trust signal into the assistant. The mechanism should subsequently be evaluated through a randomized online experiment in which participants complete repeated data-analysis tasks. By comparing trust-adaptive with non-adaptive explanation provision, the thesis should examine effects on e.g., trust calibration, reliance, and decision-making performance.

Required Skills

- Strong interest in (generative) AI and LLMs
- Good English language skills
- Basic programming skills (e.g., Python)

Starting Literature (Topic)

- Gregor, S., & Benbasat, I. (1999). Explanations from Intelligent Systems: Theoretical Foundations and Implications for Practice. *MIS Quarterly*, 23(4), 497.
- 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.
- Lee, J. D., & See, K. A. (2004). Trust in Automation: Designing for Appropriate Reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 46(1), 50–80.
- Sîrbu, A.-M., Schelhorn, T. C., Gnewuch, U. (2025) "Explanation Provision Strategies in LLM-based Data Assistants: Impact on Extraneous Cognitive Load, Trust, and Task Performance," in Proceedings of the 33rd European Conference on Information Systems (ECIS 2025).
- Srinivasan, T., & Thomason, J. (2026). Adjust for trust: Mitigating trust-induced inappropriate reliance on AI assistance. In *Proceedings of the 31st International Conference on Intelligent User Interfaces* (pp. 1883–1900).

Starting Literature (Method)

Recker, J. (2012). Research methods. In *Scientific research in information systems: A beginner's guide* (pp. 65–110). Springer. https://doi.org/10.1007/978-3-642-30048-6_5