Chair of Explainable AI-Based Business Information Systems
Prof. Dr. Ulrich Gnewuch

Bachelor's Thesis: Building “Data-Aware” LLM Analytics 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-4, Gemini, or Claude, their black-box nature often raises concerns about trustworthiness. Explainable AI (XAI) research shows that explanations can improve transparency and trust. While most explainability efforts focus on the reasoning process of the model, there is growing potential in providing users with explanations about the underlying data or datasets that the LLM assistant uses to generate its outputs. In data analysis contexts, where data is a critical input for decision-making, making such data-related aspects visible could help users better assess whether they can rely on the LLM assistant’s output.

The goal of this bachelor's thesis is to develop a “data-aware” LLM analytics assistant that provides explanations of key data-related factors, such as data quality, freshness, completeness, complexity, and source ownership, used in generating its responses. The basic assistant for performing data analysis using natural language is available, and the code will be provided to the student as a starting point. The overall development process should follow the design science research approach (Hevner et al., 2004) and include a small-scale evaluation of the prototype with fellow students.

Required Skills

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

Starting Literature (Topic)

Moges, H.-T., Vlasselaer, V. V., Lemahieu, W., & Baesens, B. (2016). Determining the use of data quality metadata (DQM) for decision making purposes and its impact on decision outcomes—An exploratory study. *Decision Support Systems*, *83*, 32–46. https://doi.org/10.1016/j.dss.2015.12.006

Price, R., & Shanks, G. (2011). The Impact of Data Quality Tags on Decision-Making Outcomes and Process. *Journal of the Association for Information Systems*, *12*(4), 323–346. https://doi.org/10.17705/1jais.00264

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). https://aisel.aisnet.org/ecis2025/hci/hci/6

Wang, R. Y., & Strong, D. M. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems*, *12*(4), 5–33. https://doi.org/10.1080/07421222.1996.11518099

Starting Literature (Method)

Vom Brocke, J., Hevner, A., & Maedche, A. (2020). Introduction to design science research. *Design science research*. *Cases*, 1-13. https://doi.org/10.1007/978-3-030-46781-4_1

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 75-105. https://doi.org/10.2307/25148625