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

Bachelor Thesis: 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-4 or Llama 3, 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 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 reduce information overload while offering necessary insights without extra user effort. However, it is not clear how to design such explanations in the LLM context.

The goal of this bachelor thesis is to develop an LLM-based data assistant displaying intelligent explanations about how it arrived at its responses. The basic assistant for performing data analysis using natural language is available, and the code can 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)

Bauer, K., von Zahn, M., & Hinz, O. (2023). Expl(AI)ned: The Impact of Explainable Artificial Intelligence on Users' Information Processing. *Information Systems Research*, *34*(4), 1582–1602. https://doi.org/10.1287/isre.2023.1199

Gregor, S., & Benbasat, I. (1999). Explanations from Intelligent Systems: Theoretical Foundations and Implications for Practice. *MIS Quarterly*, *23*(4), 497. https://doi.org/10.2307/249487

Zhao, H., Chen, H., Yang, F., Liu, N., Deng, H., Cai, H., Wang, S., Yin, D., & Du, M. (2024). Explainability for Large Language Models: A Survey. *ACM Transactions on Intelligent Systems and Technology*, *15*(2), 1–38. https://doi.org/10.1145/3639372

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