Generative and Interactive Data Storytelling for Auditing

Sara Buchmann, Emanuel Slany, Tassilo Föhr, and Stephan Scheele

Research Project Data Tales, funded by the Bavarian Collaborative Research Program (BayVFP) – Digitalization

Scan me!

Goal of the Research Project
Development of a hybrid AI approach for generative and interactive data storytelling, with applications in auditing and internal controls monitoring.

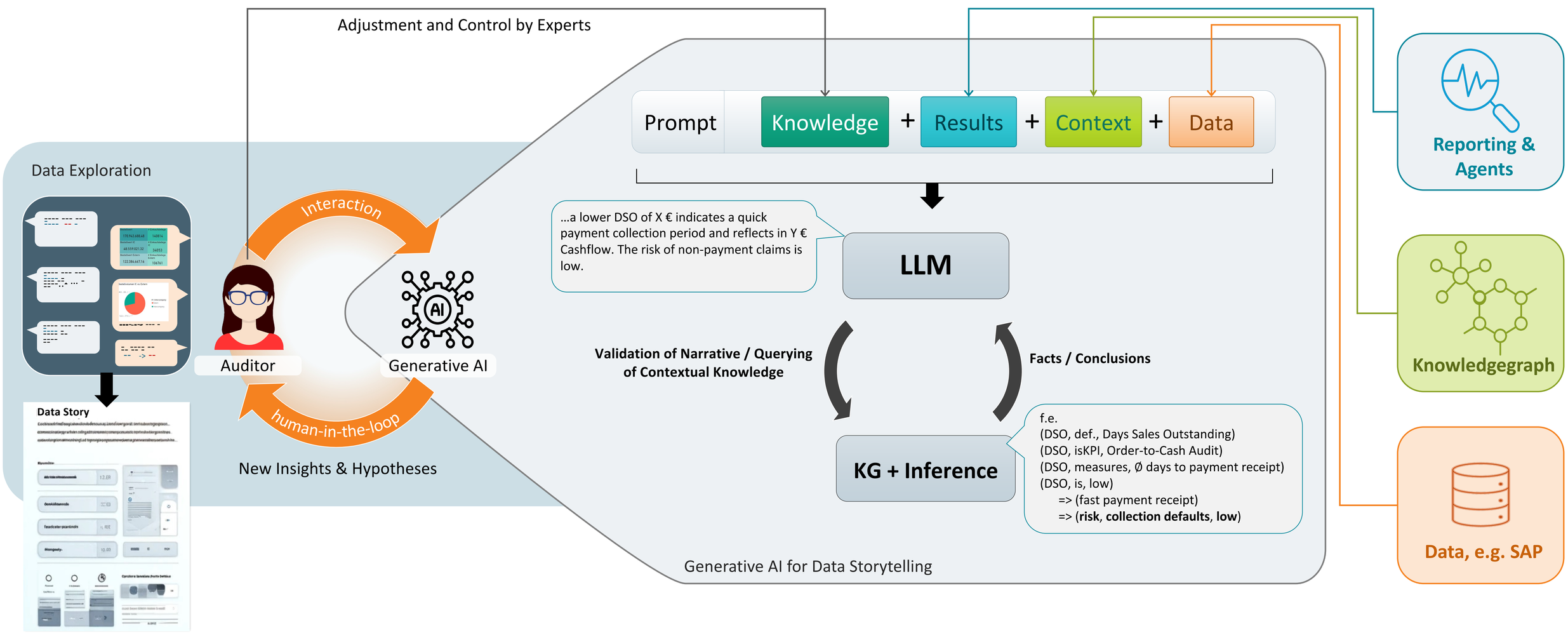
This project aims to sustainably improve the **efficiency, accuracy, and quality of audits** by empowering auditors with advanced AI-driven methodologies that enable a **deeper, content-focused exploration of audit results**.

Data Storytelling is a methodological framework that transforms complex datasets into **clear, compelling narratives**. By combining narrative techniques, data visualization, contextual knowledge, and explanations, it **makes complex analytical insights accessible and actionable** for diverse stakeholders.

Research Question
How can an AI-supported methodology be designed to **automatically generate data stories** by integrating Large Language Models with domain-specific knowledge and interactive machine learning?

Research Focus Areas

- **Hybrid AI approach:** Integration of LLMs, Machine Learning, and domain-specific knowledge to automatically create accurate and relevant data stories.
- **User-centered Human-AI collaboration:** An interactive, dialogue-based assistant refines findings through iterative user feedback, adapting to individual needs.
- **Personalization & Context Sensitivity:** Dynamically adapting stories based on user profiles, dialogue history, and query context, enabling personalized and context-sensitive storytelling that learns over time.
- **Explainability & Visualization:** Integration of XAI techniques and visual tools to improve transparency, traceability, and understanding of audit analyses.
- **Scalability & Adaptability:** Designing a flexible system scalable across various auditing applications and extendable to new domains.



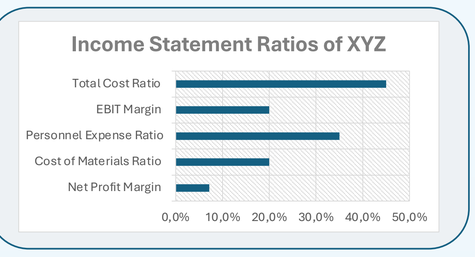
Example for an Interactive Dialogue between Auditor and Chatbot

*Hey Chatbot, I need the **key financial ratios** for this year to support the P&L Analysis of company XYZ.*

*Sure, but I'll need the **income statement data** first – including revenue, expenses, and profit figures.*
Could you specify the fiscal year and provide further details?

The relevant fiscal year is 2024.
The income statement data is uploaded.
*Can you also provide a **visualization** of the key financial ratios?*

Great, I've calculated the key ratios based on the data provided.
*Here is a **summary of the key financial ratios** for company XYZ.*



*Are there any **risks** or **irregularities**?*
*Do any of the figures stand out or require **further analysis**?*

*Yes, two figures stand out. The **personnel expense ratio** is relatively high at **35%**, and the **net profit margin** is **7.2%**, which is slightly below the typical industry benchmark of 8–10%.*

What might be driving these figures?
*Can you identify the **underlying root causes**?*

*Elevated personnel expenses may indicate **sustained cost pressure**, **increased headcount**, or **inefficiencies in workforce planning** and **utilization**.*
*The below-benchmark net profit margin of 7.2% suggests that, although revenue levels appear solid, **operating and administrative costs** are **eroding profitability** more than expected.*

*Would you like me to perform a more detailed breakdown of either the **personnel cost structure** or the **overall cost drivers** behind the net profit margin?*

*Yes, I'd like to take a **closer look** at the **net profit margin**, please.*

Combining LLM with domain-specific background knowledge and Interactive ML

We use a **Large Language Model (LLM)** to automatically generate comprehensible, narrative audit reports from complex data and to provide interactive, question-driven analytical support for auditors.

A **Knowledge Graph (KG)** is a semantic network that represents background knowledge in the form of entities, concepts, and their relations in a machine-readable way, thereby enabling logical inferences and complex queries.

We use **Interactive Machine Learning (ML)** to actively involve humans in the training process of a model through a human-in-the-loop approach, in order to continuously improve the data stories through iterative user feedback.

dab · Rödl & Partner · OTH Regensburg · VDI|VDE|IT