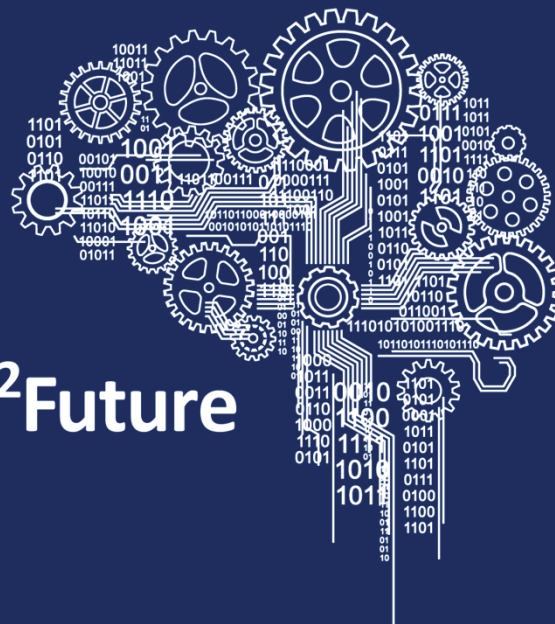


AERIAL

Retrieval-Augmented Text Generation with Large Language Models



Paul Haferlbauer¹, Matej Vukovic¹, Belgin Mutlu¹
Pro²Future GmbH¹
¹ Inffeldgasse 25F, 8010 Graz



MOTIVATION & GOALS

- Leveraging recent advancement in **Large-Language Models (LLMs)**
- **Locally** hosting an LLM and reinforcing it with **supplementary documents** (e.g., machine manuals, business reports) thereby improving its **efficiency** for **specific industrial applications**
- **Retrieval-augmented** generation to **improve** the **accuracy** and **reliability** of language models when generating responses to user queries
- Incorporating domain-specific information/knowledge into the model **reduces** its tendency towards **hallucination**, leading to more **factual responses**
- Designed to prioritize **privacy**, **security**, and **scalability**
- Support for **diverse range of use cases** beyond customer service, including **maintenance** and **onboarding**

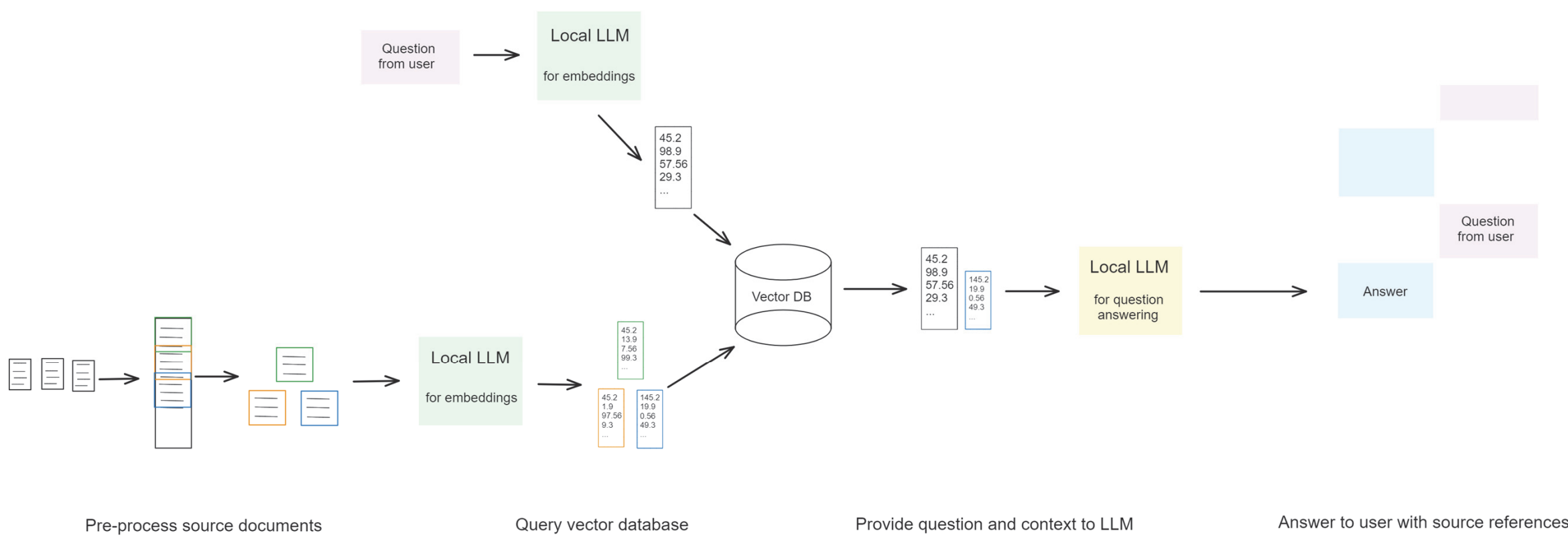
Project FactBox

Project Name UNSUDET
Project ID StratP II 3.4.2
Duration 24 Months

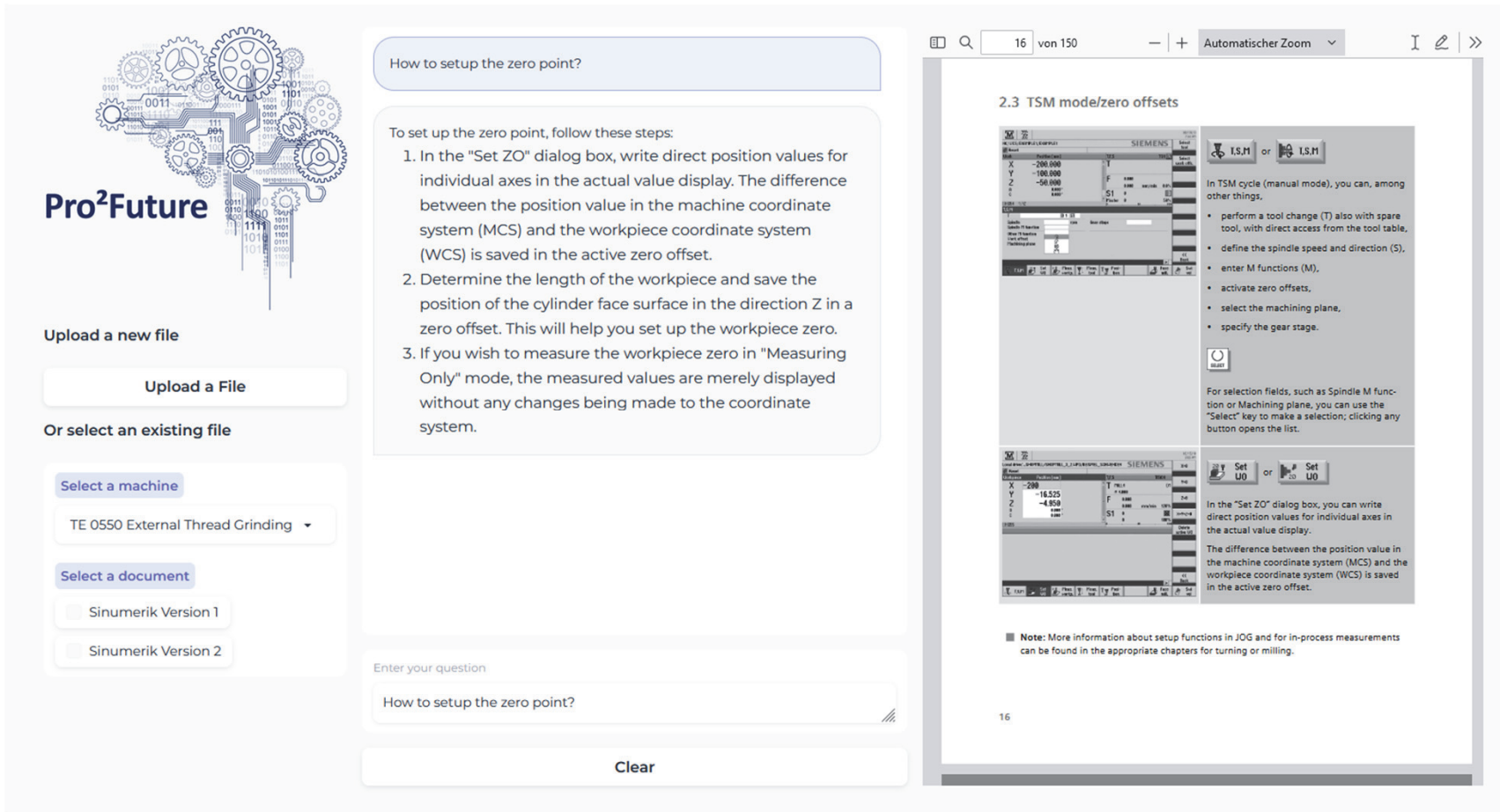
Area 3
Cognitive Decision Making

Project Lead
Dr. Belgin Mutlu

APPROACH



PROTOTYPE



FUTURE WORK

- Enhancing model **transparency** through knowledge graphs
- Context- and user aware **prompt recommendations**
- **Multi-modal responses** for enhanced user interaction (visual vs. text-based responses)
- Utilizing similarity scores for documents and group of texts as quantifiable measures of **certainty** and **accuracy**
- Integration of **causal knowledge** into retrieval augmented generation

CONTRIBUTION

Economic & Scientific contribution

- **Improved Operational Efficiency:** Locally hosting and reinforcing LLMs with industry-specific documents leads to faster and more accurate responses to customer queries and internal processes
- **Reduced Training Costs:** Using existing pre-trained models reduces the costs associated with training language models, including computational resources
- **Knowledge Integration:** Incorporating domain-specific knowledge into LLMs can reduce the risk of generating incorrect or false information
- **Minimized Data Privacy Risks:** Local hosting increases privacy and security

Contact: DI Paul Haferlbauer, Pro²Future GmbH, paul.haferlbauer@pro2future.at
Acknowledgement: This work was supported by Pro²Future (FFG, 881844)