



Pro²Future

Trustworthy Digital Systems Assisted by Large Language Models (LLMs) - TrustinLLM

21st of October 2025

DI Richard Hohensinner, Dr. Belgin Mutlu



Shareholders of Pro2Future GmbH:



Public funding of Pro²Future:

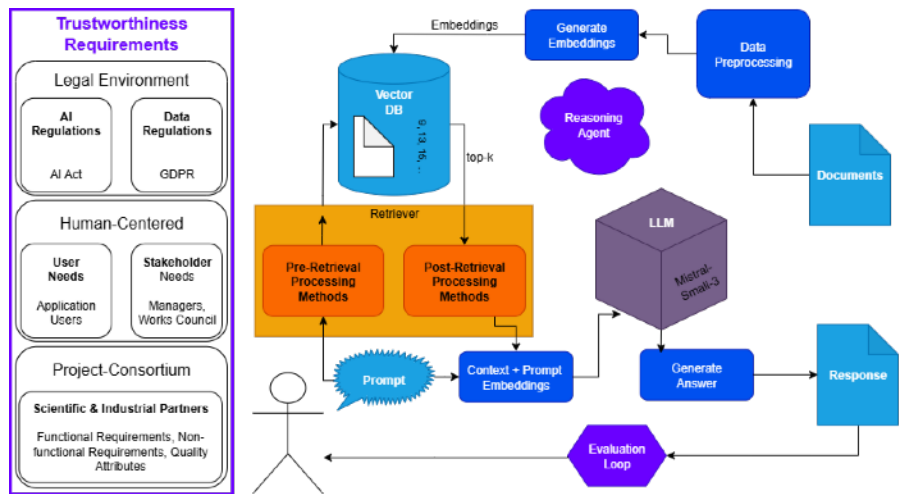


Problem Statement & Motivation

- Modern society is increasingly reliant on complex digital systems across all sectors
- These systems are rapidly integrating AI technologies, e.g., **Large Language Models (LLMs)**
- This deep integration raises critical concerns around:
 - **Trustworthiness** : ensuring systems behave as expected under all conditions.
 - **Transparency** : making AI decisions understandable and explainable.
 - **Reliability** : maintaining consistent and robust performance.
 - **Safety** : preventing harmful or incorrect outputs.
- Motivation for this project is to
 - Research technologies that improve the **trustworthiness** of AI-integrated systems
 - Support the development of systems capable of generating **accurate** and safe assistance/responses
 - Enable the creation of digital infrastructures that are **transparent, dependable, and secure**.

Approach :: Trustworthy AI - Retrieval Augmented Generation (RAG)

- **Utilize Retrieval-Augmented Generation (RAG):**
 - Ground AI-generated responses in authoritative source documents to reduce hallucinations.
 - Improve factual accuracy by anchoring outputs in verified content.
- **Enhance RAG with Pre- and Post-Retrieval Processing:**
 - Pre-retrieval: Optimize query formulation to improve document relevance.
 - Post-retrieval: Refine and validate generated responses using retrieved content.
- **Implement Transparency and Traceability Measures:**
 - Track and expose the origin of information used in responses.
 - Enable auditability of AI decisions and outputs.
- **Design Human-Centered User Interfaces:**
 - Clearly communicate uncertainties and potential hallucinations in AI responses.
 - Provide intuitive visual cues and explanations to support user understanding and trust.



TrustinLLM - Contributions

SCIENTIFIC

- Increasing the **transparency** capabilities of **LLM-based RAG systems**
- Increased **trustworthiness** for AI-based systems
- **Data provenance** and **traceability** solutions to tackle rising data obfuscation
- **Mitigation** of hallucinations
- **Multi-modal** reasoning

BUSINESS

- Lower risk of costly errors through **reliable** AI outputs
- Higher **productivity** via reduced manual oversight
- Faster adoption due to increased user trust
- Improved compliance with **safety** and **transparency**
- Competitive advantage through **trustworthy** AI products

Thank you!

Questions welcome!

Contact



DI. Richard Hohensinner

PhD Student

Area Analytics

Richard.hohensinner@pro2future.at



Dr. Belgin Mutlu

Area Manager, Key Researcher

Area Analytics

belgin.mutlu@pro2future.at

+43 664 88371 323



Shareholders of Pro2Future GmbH:



Public funding of Pro²Future:





Pro²Future

Cognitive and Sustainable Products and Production Systems of the Future!

Shareholders of Pro2Future GmbH:



Public funding of Pro²Future:

