Workshop

Grounding Generative AI Models

Holger Keibel, Johannes Porzelt

Karakun AG

Description

One fundamental shortcoming of generative AI systems such as ChatGPT is that they tend to produce hallucinatory outputs which the human user might easily take to be facts, with potentially disastrous consequences. In principle, it is the responsibility of the user to fact-check any generative AI output before relying on it, but this is not common practice and often impossible for users when the respective topic lies outside their field of expertise.

This calls for (semi-)automated approaches to grounding generative AI models in relevant facts – or at least to quantify the confidence that a given generative AI output is correct. Existing approaches range from integrating domain-specific knowledge directly into the models (typically by means of retrieval-augmented generation, RAG) to applying post-generation filtering techniques (automated fact-checking) to making the generation of the AI output transparent to the user (explainable AI).

This workshop aims at sharing practical experience applying such approaches. The three presentations apply LLMs generatively on domains ranging from ticket routing in customer support to teaching assistants to the ethically challenging domain of euthanasia decisions. Hallucinations of the generative AI approaches are problematic in all three domains, and the presentations describe different approaches to mitigating them: RAG against a fixed set of outputs in conjunction with supervised learning approaches, RAG against multimodal Knowledge Graphs (KGs), and an explainable AI approach by querying the LLM itself in a series of yes/no questions. In a final panel discussion, lessons learned will be shared and discussed.

Schedule

10:30 Holger Keibel & Johannes Porzelt (Karakun AG, Basel) Introduction

10:45 Flurin Gishamer (Open Systems, Zurich) Practical Strategies for Enhancing Reliability of GenAI Systems in Customer Operations: An Overview

11:10 Xiaokun Zhang (Athabasca University, Edmonton/Canada) Knowledge Graphs Enhanced Retrieval-Augmented Generation for Eliciting Higher-Order Thinking

11:35 Gerold Schneider (UZH, Zurich) Evaluating Transformers on the Ethical Question of Euthanasia

12:00 All speakers Panel Discussion: Lessons Learned, Q & A
12:30 End of workshop, lunch break