Enhancing Data Precision with Large Language Models: Analyzing Failures and Innovating Database Curation

Georg Gottlob[1]

[1]*University of Calabria, Italy*

Abstract

On 25th June 2024, Georg Gottlob delivered a keynote talk at the 32nd Symposium on Advanced Database Systems in Villasimius (Sardinia, Italy). The following is the abstract of his talk and a short biography

Abstract of the Keynote

The advent of Large Language Models (LLMs) such as ChatGPT represents a significant milestone in the AI revolution. This talk commences with an exploration of text-based generative AI tools, highlighting exemplary performances in producing elegantly crafted texts. However, LLMs often fail, particularly when tasked with generating precise data absent from established databases like Wikipedia. This phenomenon is critically examined through a "psychoanalysis" of LLMs that identifies fundamental causes for such failures and hallucinations. In response to these challenges, the second part of the talk introduces the Chat2Data method and system, an innovative framework designed to harness the capabilities of LLMs for the automatic generation, enrichment, and verification of databases and data sets. Chat2Data automatically generates sophisticated workflows that incorporate problem decomposition, strategic LLM querying, and meticulous analysis of responses. To refine reliability and accuracy, the system integrates supplementary technologies such as Retrieval-Augmented Generation (RAG), rule-based knowledge processors, and data-graph analysis. This comprehensive approach not only mitigates the pitfalls identified but also significantly advances the utility of LLMs in complex data environments.

Short Biography

Georg Gottlob is a Professor of Computer Science at the University of Calabria. Until recently, he was a Royal Society Research Professor at the Computer Science Department of the University of Oxford, a Fellow of St John's College, Oxford, and an Adjunct Professor at TU Wien. His interests include knowledge representation, database theory, query processing, web data extraction, and (hyper)graph decomposition techniques. Gottlob has received the Wittgenstein Award from the

---

Austrian National Science Fund and the Ada Lovelace Medal (UK). He is a Fellow of the Royal Society, and a member of the Austrian Academy of Sciences, the German National Academy of Sciences, and the Academia Europaea. He was a founder of Lixto, a web data extraction firm acquired in 2013 by McKinsey & Company. In 2015 he co-founded Wrapidity, a spin out of Oxford University based on fully automated web data extraction technology developed in the context of an ERC Advanced Grant. Wrapidity was acquired by Meltwater, an internationally operating media intelligence company. Gottlob then co-founded the Oxford spin-out DeepReason.AI, which provided knowledge graph and rule-based reasoning software to customers in various industries. DeeoReason.AI was also acquired by Meltwater.