Implementation of Fact-checking pipelines for LLMs
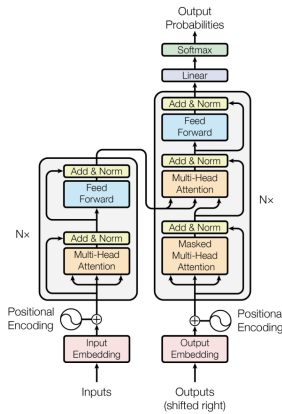
Figure 1: The Transformer - model architecture.
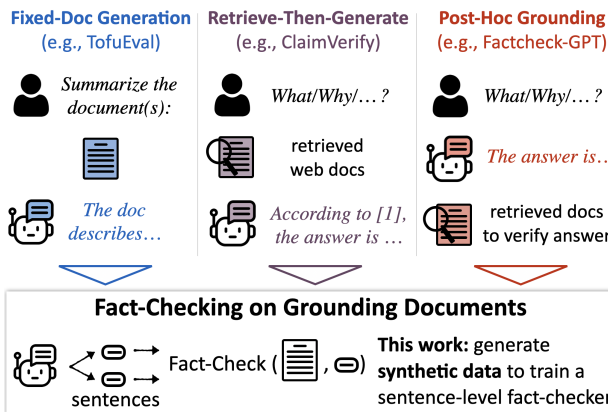
Figure 1: (Left) Transformer architecture [2]. **(Right)** Fact-checking pipeline from [1].

Background

Large language models (LLMs) like OpenAI's GPT or Google's Bard have become increasingly influential due to their ability to generate human-like text. However, they often produce inaccurate or misleading information—a phenomenon referred to as "hallucination". This limitation raises concerns about their use in critical applications, such as journalism, education, healthcare, and law. To mitigate these issues, there is a growing interest in implementing fact-checking pipelines that evaluate and improve the factual accuracy of LLM outputs.

A fact-checking pipeline integrates retrieval systems, external databases, and logical reasoning to validate information generated by the LLM. By leveraging state-of-the-art tools like web scraping, natural language processing (NLP), and knowledge graphs, these pipelines can detect inaccuracies and provide reliable corrections.

Challenges in Fact-Checking pipelines:

- Distinguishing factual claims from general text.
- Retrieving contextually relevant and authoritative evidence.
- Verifying multi-faceted claims and resolving ambiguity.
- Presenting corrections without disrupting the user experience.

Description

The motivated student will gather information about state-of-the-art approaches for LLMs and fact-checking pipelines. Promising pipelines will be implemented and evaluated qualitatively.

During this project the student will be

- working with state-of-the-art LLMs and NLP approaches,
- understanding the theory of LLMs,
- exploring challenges of fact-checking pipelines,
- implementing existing solutions for own problems.

Technical University of Munich

TUM School of Computation, Information and Technology (CIT)

Chair of Robotics, Artificial Intelligence and Real-time Systems

Supervisor:
Prof. Dr.-Ing. Alois Knoll

Advisor:
Nico Reeb, M.Sc.

For more information please contact us:

Phone: +49.89.289.17629

E-Mail: nico.reeb@tum.de

Internet: www.ce.cit.tum.de/air

References

[1] Liyan Tang, Philippe Laban, and Greg Durrett. "MiniCheck: Efficient Fact-Checking of LLMs on Grounding Documents". In: *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*. Ed. by Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen. Miami, Florida, USA: Association for Computational Linguistics, Nov. 2024, pp. 8818–8847. URL: https://aclanthology.org/2024.emnlp-main.499.

[2] Ashish Vaswani et al. *Attention Is All You Need*. 2023. arXiv: 1706.03762 [cs.CL]. URL: https://arxiv.org/abs/1706.03762.

Technical University of Munich

TUM School of Computation, Information and Technology (CIT)

Chair of Robotics, Artificial Intelligence and Real-time Systems