Title: Hallucination to Truth: A Review of Fact-Checking and Factuality Evaluation in Large Language Models

Author names: Subhey Sadi Rahman, Md. Adnanul Islam, Md. Mahbub Alam, Musarrat Zeba, Md. Abdur Rahman, Sadia Sultana Chowa, Mohaimenul Azam Khan Raiaan, Sami Azam

Abstract: Large Language Models (LLMs) are trained on vast and diverse internet corpora that often include inaccurate or misleading content. Consequently, LLMs can generate misinformation, making robust fact-checking essential. This review systematically analyzes how LLM-generated content is evaluated for factual accuracy by exploring key challenges such as hallucinations, dataset limitations, and the reliability of evaluation metrics. The review emphasizes the need for strong fact-checking frameworks that integrate advanced prompting strategies, domain-specific fine-tuning, and retrieval-augmented generation (RAG) methods. It proposes five research questions that guide the analysis of the recent literature from 2020 to 2025, focusing on evaluation methods and mitigation techniques. Instruction tuning, multi-agent reasoning, and RAG frameworks for external knowledge access are also reviewed. The key findings demonstrate the limitations of current metrics, the importance of validated external evidence, and the improvement of factual consistency through domain-specific customization. The review underscores the importance of building more accurate, understandable, and context-aware fact-checking. These insights contribute to the advancement of research toward more trustworthy models.

https://doi.org/10.1007/s10462-025-11454-w

https://arxiv.org/abs/2508.03860