Large language models for writing scientific reviews

Andrey Guskov, PhD.

HEAD OF THE LABORATORY OF SCIENTOMETRICS AND SCIENCE COMMUNICATION
RUSSIAN RESEARCH INSTITUTE OF ECONOMICS, POLITICS AND LAW IN SCIENCE AND TECHNOLOGY

GUSKOV.ANDREY@GMAIL.COM
Using LLM* gives an advantage

The quality of work of employees using AI in analytical tasks is 40% higher.

*LLM – Large Language Models

https://dx.doi.org/10.2139/ssrn.4573321
At least 30% of researchers used LLM in 2023

- Additional source of ideas
- It’s easier to work with a finished draft
- Improving Academic Writing

“The goal of a researcher is to do science, not write papers”.

https://www.nature.com/articles/d41586-023-02980-0
Perplexity.ai

AI chatbot-powered research and conversational search engine that answers queries using natural language predictive text.

https://www.perplexity.ai
ChatDOC
Chat with papers

https://chatdoc.com/
YouTube-channel about academic AI-tools

https://www.youtube.com/@DrAndyStapleton
Writing reviews
The first “AI-written” scientific review was published in 2019

Technical workflow described in chapter 1
Stages of creating a scientific review

- **Goal-setting**: Identifying research questions for review
- **Searching**: Searching for publications according to formal criteria (WoS, Scopus, GS…)
- **Screening**: Primary selection of relevant publications
- **Archiving**: Formation of a full-text library for review
- **Retrieving**: Extracting answers to research questions from full texts
- **Summarizing**: Composing answers into review chapters
- **Assembling**: Creating draft of the review
- **Editing**: Expert verification and text revision
Typical questions that authors deal with

- Is the publication fit for the topic of the review?
- Which results of the paper correspond to the research questions?
- How to summarize selected results?
- What are the promising areas of research?
Act as an experienced medical researcher. Your task is to determine the relevance of a given paper to a specific research question. Based on the title and abstract of the paper, assess whether the paper contains research results that directly or partially answer the research question, or if it provides methods or data that could help investigate the question further.

Instructions:
1. Read the Title and Abstract: Carefully review the title and abstract of the paper to understand its focus, objectives, and scope.
2. Analyze for Relevance: Evaluate whether the paper addresses the research question directly, provides partial answers, or offers methods or data useful for further investigation of the question.
3. Determine Relevance: Decide if the paper should be included in the literature review based on its relevance to the research question.

Research question:%QUESTION%
Paper title: %TITLE%
Paper abstract: %ABSTRACT%
Output: RELEVANT or NOT RELEVANT
Question 2. Which results of the paper correspond to the research questions?
Question 3. How to summarize selected results?

Question 4. What are the promising areas of research?

Types of summarizers

- Summarize the results...
- Propose promising areas of research...
- ...

How the motivation of professional growth affects scientific mobility?

Researchers in the field of artificial intelligence tend to move early in their careers, particularly through a type of mobility called academic mobility. This type of mobility requires researchers to accumulate more research publications and establish collaborations with corporate researchers and high-impact collaborators. The findings suggest that the motivation for professional growth influences scientific mobility, as researchers engaging in academic mobility need to focus on building research and social capital to enhance their success. This research provides insights for researchers and policymakers on career development and talent recruitment in the field of artificial intelligence. [Chen2023.pdf]

Scientific mobility positively impacts performance and success regardless of career stage, with researchers initiating mobility later experiencing more significant benefits. The study shows that mobility enhances scientific outcomes, as evidenced by increased publications per year and citations per publication. Timing is crucial, with advanced career-stage mobility leading to more positive results. This highlights the importance of motivation for professional growth in driving scientific mobility’s effectiveness on researchers’ achievements. [Dianese2022.pdf]

Italian researchers’ motivation for professional growth significantly affects scientific mobility. The study reveals that dissatisfaction with career prospects within the Italian Research and Innovation (R&I) system drives Italian researchers to emigrate and pursue academic careers abroad. Researchers in Italy report worse working conditions, including limited permanent contracts and dissatisfaction with remuneration. In contrast, Italian researchers working abroad experience faster career progression, transparency in recruitment, and higher confidence in future career prospects. The study emphasizes the need for policy actions to improve Italy’s research and university system, including increased resources, recruitment, and better remuneration. [Dianese2021.pdf]

How the socio-economic reasons affect scientific mobility?

Socio-economic reasons have a modest impact on scientific mobility, as scientists who move from elite institutions to lower-ranked ones experience a slight decrease in scientific performance. However, transitioning into elite institutions does not lead to subsequent performance gain. This conclusion is based on the analysis of over 420,000 papers and the affiliation information of individual scientists, allowing for the reconstruction of their career trajectories. The study also highlights the importance of institutional ranking and suggests the need for further investigations using more comprehensive data sources. [Dianese2016.pdf]
Stages 1-4. Searching, screening and collecting full-texts
Stages 5-8. Processing full-texts
Stages 5-8. Processing full-texts
Stages 5-8. Processing full-texts
Check: reverse tracking

Check can also be AI-automized
Fact-checking
Frameworks for Fact-checking

FACTSCORE: Fine-grained Atomic Evaluation of Factual Precision in Long Form Text Generation

Seonwoo Shin1, Kalpesh Krishna1, Xiuxi Lian2, Mike Lewis1, Wen-tau Yih1
Pang Wei Koh1, Mohit Iyyer3, Luke Zettlemoyer1, Hannaneh Hajishirzi1,2
1University of Washington, 2University of Massachusetts Amherst
https://arxiv.org/abs/2305.14251

Abstract

Evaluating the factual accuracy of long form text generated by large language models (LLMs) is non-trivial because (1) LLMs generate a mixture of supported and unsupported pieces of information, making binary judgments of quality challenging; and (2) human evaluation is time-consuming and costly. In this paper, we introduce FACTSCORE, a new evaluation that breaks generation into a series of atomic facts and computes the percentage of atomic facts supported by a suitable knowledge source. We conduct an extensive human evaluation to obtain FACTSCOREs of a subset of paragraphs generated by several state-of-the-art commercial LLMs—GPT-4, ChatGPT, and the removal-argumented FactualGPT—and report new analysis demonstrating the need for such a fine-grained score (e.g., ChatGPT only achieves 56%).

Introduction

Long-form factuality in large language models

Jerry Wei1,2, Chengwen Yang1, Xingyi Song3, Yifeng Lu4, Da Hu1
Jie Huang1,2, Daxin Tran3, Ruiyu Liu1, Nathan Jin1,2, Qiao V. Le1
1Google DeepMind, 2Stanford University, 3University of Illinois at Urbana-Champaign

Long-form factuality in large language models

Large language models (LLMs) often generate content that contains factual errors when responding to fact-seeking prompts on open-ended topics. To benchmark a model’s long-form factuality in open domains, we first use GPT-4 to generate LongFact, a prompt set comprising thousands of questions spanning 38 topics. We then propose that LLM agents can be used as automated evaluators for long-form factuality through a method in which we call Search-Augmented Factuality Evaluator (SAFE). SAFE utilizes an LLM to break down a long-form response into a set of individual facts and to evaluate the accuracy of each fact using a multi-step reasoning process comprised of sending search queries to Google Search and determining whether a fact is supported by the search results. Furthermore, we propose extending F1 score as an aggregated metric for long-form factuality. To do so, we balance the percentage of supported facts in a response (precision) with the percentage of provided facts relative to a hyperparameter representing a user’s preferred recall level (recall).

Empirically, we demonstrate that LLM agents can outperform crowdsourced human annotators—on a set of ~10k individual facts, SAFE agree with crowdsourced human annotators 72% of the time, and in a random subset of 100 disagreement cases, SAFE wins 76% of the time. At the same time, SAFE is more than 20 times cheaper than human annotators. We also benchmark diverse language models on LongFact across four model families (GPT, QPF, Clude, and Dalm-M2), finding that larger language models generally achieve better long-form factuality. LongFact, SAFE, and all experimental code are available at https://github.com/google-deepmind/long-form-factuality.
Idea 1: Divide and check!

Divide text (using LLM) into atomic facts and evaluate them relative to a given source of knowledge.
Idea 2: Assessing “factuality” using search engines

![Search-Augmented Factuality Evaluator (SAFE)](https://arxiv.org/abs/2403.18802)
LLM agents do fact-checking better and cheaper than people

The discrepancy between SAFE and humans is 28%.

In case of discrepancies, SAFE wins much more often.

SAFE is 20 times cheaper.
Conclusion

1. LLMs will be used in research more and more frequently.
2. This will put a new burden on the system of scientific communication and provide new tools to cope with it.
3. To implement new tools, scarce competencies are needed.
Large language models for writing scientific reviews

Andrey Guskov, PhD.

HEAD OF THE LABORATORY OF SCIENTOMETRICS AND SCIENCE COMMUNICATION

RUSSIAN RESEARCH INSTITUTE OF ECONOMICS, POLITICS AND LAW IN SCIENCE AND TECHNOLOGY

GUSKOV.ANDREY@GMAIL.COM