Delving into ChatGPT usage in academic writing through excess vocabulary

Dmitry Kobak[1,2], Rita González-Márquez[1,2], Emőke-Agnes Horvát[3] & Jab Lause[1,2]

[1]*Hertie Institute for AI in Brain Health, Universität Tübingen,* [2]*Tübingen AI Center,* [3]*Northwestern University*

dmitry.kobak@uni-tuebingen.de, rita.gonzalez-marquez@uni-tuebingen.de, a-horvat@northwestern.edu, jan.lause@uni-tuebingen.de

Recent large language models (LLMs) can generate and revise text with human-level performance, and have been widely commercialized in systems like ChatGPT. These models come with clear limitations: they can produce inaccurate information, reinforce existing biases, and be easily misused. Yet, many scientists have been using them to assist their scholarly writing. How wide-spread is LLM usage in the academic literature currently? To answer this question, we use an unbiased, large-scale approach, free from any assumptions on academic LLM usage. We study vocabulary changes in 14 million PubMed abstracts from 2010–2024, and show how the appearance of LLMs led to an abrupt increase in the frequency of certain style words. Our analysis based on excess words usage suggests that at least 10% of 2024 abstracts were processed with LLMs. This lower bound differed across disciplines, countries, and journals, and was as high as 30% for some PubMed sub-corpora. We show that the appearance of LLM-based writing assistants has had an unprecedented impact in the scientific literature, surpassing the effect of major world events such as the Covid pandemic.