Can AI reproduce disciplinary voice?

A comparative corpus-based study of GPT-4's ability to reproduce disciplinary voice in AI-generated linguistic prose

EA LINDHARDT OVERGAARD & ULF DALVAD BERTHELSEN

The purpose of this article is to uncover the extent to which generative AI models – particularly GPT-4 – can reproduce disciplinary voice in Danish academic prose. These new large language models come with a promise of transforming our writing practices, including academic writing. However, the quality of the autogenerated contributions, especially when the models are used for smaller languages like Danish, is still unclear. We focus on disciplinary voice because it is a relatively well-described phenomenon that can also be examined quantitatively through the analysis of the surface structure of texts. We specifically focus on three aspects of disciplinary voice, *stance, engagement*, and *discipline-specific vocabulary*, investigating them quantitatively through a corpus-based comparative study. We compare a corpus of Danish linguistic articles with a corpus of AI-generated academic prose with linguistic content. Our study indicates that AI-generated texts significantly differ from the authentic texts in some areas. In the discipline-specific vocabulary category, the differences are notably large, whereas the stance and engagement categories exhibit smaller differences. In these two latter categories, the differences are so marginal that AI-generated texts appear to reproduce the disciplinary voice in a manner that, from a quantitative perspective, mirrors what we observe in authentic texts.

KEYWORDS: generative artificial intelligence; chatbots; text quality; disciplinary voice; corpus linguistics