Prompted generative artificial intelligence versus interdisciplinary clinicians for academic abstract writing: the AbChat Collaborative short protocol.

Introduction

In the era of artificial intelligence (AI), large language models (LLMs) are increasingly being used within medicine and academia. Whilst many have shown that generative AI can summarise entire research studies efficiently, research is often ready for presentation long before it is transcribed into a full manuscript for journal submission.

Aims

This study will determine the ability of a leading LLM (chatGPT-4-turbo) to generate academic conference abstracts using pre-specified prompts and compare these to the same abstracts as written by clinicians from a variety of medical fields, with the goal of validating LLMs as an academic tool.

Study Design

Abstracts submitted by clinicians in a particular field, previously accepted for conference presentation, will be summarised into approximately 100-word bullet point summary prompts. The prompts will be provided to chatGPT-4-turbo via a Python API to generate a 300-word abstract, to be directly compared against the original abstract. Abstracts will be assigned a random code using inbuilt python functionality. Four independent, blinded, senior academic adjudicators from each specialty will score a random selection of AI or clinician abstracts according to a previously validated proforma. The adjudicators will additionally be asked whether they think the abstract was written by clinicians or generative AI. Four adjudicators will each score n=47 abstracts, a mixture of AI or clinician generated, such that each human and AI abstract receives a total score from two raters. Each individual project will use the same four adjudicators and this methodology will be followed for each medical discipline included within the study.

Outcomes

The primary outcome will be the abstract score for the generative AI abstracts versus the clinician abstracts. The secondary outcomes will be accuracy of the generated abstract as compared to the original prompt, the performance of the LLM as an abstract scorer versus field experts (including the intra-class correlation coefficients (ICC)) and the percentage of plagiarism (using an online free plagiarism checker) as well as originality (scored using an AI output detector).

Roles & responsibilities

The AbChat Collaborative is headquartered in Charing Cross Hospital, London, UK. It is the coordinating centre, run by the chief investigator Dr Benedict Turner and the AI committee. The Scientific Committee comprises all the principal investigators for each study. The Members of the Scientific Committee will approve the study design and protocol of the AbChat Collaborative. The AI Committee will generate the AI abstracts using the specified Python code and chatGPT-4-turbo API. A complete Team Member list will be made available online.

Authorship

All who partake in the study will hold collaborative authorship status on every publication. Additionally, principal investigators will be primary authors on their specific publication, whilst data monitors and senior assessors will also hold named

authorship on their given study. The chief and co-chief investigators will be senior authors on all published articles.

Medical disciplines
Principal authors that are current specialty trainees will be recruited from all 26 disciplines of medicine and will all liaise directly with the single central AI committee

Power calculations
With an effect size of 68%, sigma of 1, alpha of 0.05 and beta of 0.1 for a 90% power to detect an effect, a total of 94 abstracts (47 in each group) are required per study.

Data ownership
The AbChat Collaborative will act as the custodian of the data. All participants will be able to access their own submitted data without the need for permission from the AbChat Collaborative.

Data confidentiality
There will be no individual or centre-related information included within the abstracts, all data will be fully anonymized.

Timeline
From the go-live date for each individual project, the turnaround time for completion of data collection and abstract scoring is 8 weeks. Principal investigators will be contacted once per fortnight by the chief investigators to check for any problems in establishing the project. A single extension of 2 weeks may be applied to any individual study at the discretion of the chief investigators, after which individual projects may be terminated and a new principal investigator sought.