GPT For RCT

Wrightson, J.G., Blazey, P., Moher, D. Khan, K.M., Ardern, C.L.

### Background

Poor reporting of clinical trials is common 1, threatens the reliability and credibility of medical research 2 and affects patient care 3. Improving trial reporting, therefore, is an ethical imperative 4,5. Using reporting guidelines, such as the CONsolidated Standards of Reporting Trials (CONSORT), can improve trial reporting standards 6–8, but adherence is often poor 9. We have recently shown 10 that it is possible to use an Artificial Intelligence Large Language Model (LLM) to determine how well a clinical trial report adheres to reporting guidelines. In this study, we will expand this work and develop and evaluate an open-source LLM (GPT\_for\_RCT) to allow authors, publishers, peer reviewers and other research funders to quickly and accurately check the reporting standards of clinical trial reports.

### Approach

This study has two parts. Initially, we will use the GPT-4 (OpenAI, USA) LLM to build a corpus of 1000 synthesized clinical trial reports. This dataset will be used to train the GPT\_for\_RCT LLM. Each report in the corpus will include labels that indicate which CONSORT reporting guidelines they do and do not adhere to. This dataset will be published in a peer-reviewed journal and made accessible to other researchers to use and evaluate. In the second part of the study, we will train a number of open-source LLMs and the closed-source GPT-4 model on the synthesized text corpus. We will compare how well each model evaluates adherence to CONSORT reporting guidelines by comparing their performance to published datasets wherein experts evaluated samples of clinical trial reports for CONSORT adherence. We will select the most accurate open-source LLM (defined as the model with performance metrics as close or equal to the performance of humans and the GPT-4 LLM). We will make the GPT\_for\_RCT LLM open-source and available to other researchers.

### Requirements

We need assistance with part one of this study: the creation and validation of the synthesized text corpus. We have to check that the text the LLM produces for each paper is representative of the text found in clinical trial reports and that the labels the model provides (i.e. the reporting guidelines the model states the text does and does not adhere to) are accurate. We expect this process to be iterative and to require “prompt engineering’, which is the systematic evaluation of model output based on changes to the instructions given to the model.

### Expected Outcomes and Benefits

We will provide training in evaluating clinical trial reports and, if you are interested, in prompt engineering techniques and coding practices (though this is optional). We expect to publish the evaluation of the dataset and the dataset itself in a peer-reviewed journal (e.g. <https://www.nature.com/sdata/journal-information>), and you would have the opportunity to contribute to these publications as an author (please see the ICMJE guidelines ([LINK](https://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html)) for details related to authorship of peer-reviewed journal articles).

References

1. Dechartres A, Trinquart L, Atal I, et al. Evolution of poor reporting and inadequate methods over time in 20 920 randomised controlled trials included in Cochrane reviews: research on research study. *BMJ*. 2017;357:j2490. doi:10.1136/bmj.j2490

2. Simera I, Moher D, Hirst A, Hoey J, Schulz KF, Altman DG. Transparent and accurate reporting increases reliability, utility, and impact of your research: reporting guidelines and the EQUATOR Network. *BMC Med*. 2010;8(1):24. doi:10.1186/1741-7015-8-24

3. Duff JM, Leather H, Walden EO, LaPlant KD, George TJ. Adequacy of Published Oncology Randomized Controlled Trials to Provide Therapeutic Details Needed for Clinical Application. *JNCI J Natl Cancer Inst*. 2010;102(10):702-705. doi:10.1093/jnci/djq117

4. Chalmers I. Underreporting research is scientific misconduct. *JAMA*. 1990;263(10):1405-1408. doi:10.1001/jama.1990.03440100121018

5. Chalmers I, Bracken MB, Djulbegovic B, et al. How to increase value and reduce waste when research priorities are set. *The Lancet*. 2014;383(9912):156-165. doi:10.1016/S0140-6736(13)62229-1

6. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMC Med*. 2010;8:18. doi:10.1186/1741-7015-8-18

7. Pandis N, Shamseer L, Kokich VG, Fleming PS, Moher D. Active implementation strategy of CONSORT adherence by a dental specialty journal improved randomized clinical trial reporting. *J Clin Epidemiol*. 2014;67(9):1044-1048. doi:10.1016/j.jclinepi.2014.04.001

8. Hopewell S, Ravaud P, Baron G, Boutron I. Effect of editors’ implementation of CONSORT guidelines on the reporting of abstracts in high impact medical journals: interrupted time series analysis. *BMJ*. 2012;344:e4178. doi:10.1136/bmj.e4178

9. Samaan Z, Mbuagbaw L, Kosa D, et al. A systematic scoping review of adherence to reporting guidelines in health care literature. *J Multidiscip Healthc*. 2013;6:169-188. doi:10.2147/JMDH.S43952

10. Wrightson JG, Blazey P, Moher D, Khan KM, Ardern CL. GPT for RCTs? Using AI to determine adherence to reporting guidelines. Published online March 24, 2024:2023.12.14.23299971. doi:10.1101/2023.12.14.23299971