RCTCheck Report

Instructions: This automated report compares your [clinical trial protocol / clinical trial report] with the [SPIRIT / CONSORT] guidelines. It provides some suggestions to improve your report based on the probability that recommended [SPIRIT / CONSORT] items are included in your manuscript. The report is not intended to be a substitute for a thorough review of the [SPIRIT / CONSORT] guidelines. Please review the report and make any necessary changes to your manuscript. If you have any questions, please contact us at email@email.com

1. Some items might not be reported, or might not be reported clearly

Consider adding information about the following items. If these items are reported in your manuscript, consider clarifying them by using more common words, phrases, and headings such as those in the [SPIRIT / CONSORT] explanatory guidance.

1a Identification as a randomized trial in the title
No corresponding text identified

1b Structured summary of trial design, methods, results, and conclusions
No corresponding text identified

2a Scientific background and explanation of the rationale
No corresponding text identified

2b Specific objectives or hypotheses
No corresponding text identified

3a Description of trial design (such as parallel, factorial) including allocation ratio
No corresponding text identified

3b Important changes to methods after trial commencement (such as eligibility criteria), with reasons
No corresponding text identified

4a Eligibility criteria for participants
No corresponding text identified

4b Settings and locations where the data were collected
No corresponding text identified

5 The interventions for each group with sufficient details to allow replication, including how and when they were actually administered
No corresponding text identified

6a Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed
No corresponding text identified

6b Any changes to trial outcomes after the trial commenced, with reasons
No corresponding text identified

7a How sample size was determined?
No corresponding text identified

7b When applicable, explanation of any interim analyses and stopping guidelines
No corresponding text identified

8a The method used to generate the random allocation sequence
No corresponding text identified

8b Type of randomization; details of any restriction (such as blocking and block size)
No corresponding text identified

9 The mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned
No corresponding text identified

10 Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions
No corresponding text identified

11a If done, who was blinded after assignment to interventions (e.g., participants, care providers, those assessing outcomes) and how
No corresponding text identified

11b If relevant, description of the similarity of interventions
No corresponding text identified

12a Statistical methods used to compare groups for primary and secondary outcomes
No corresponding text identified

12b Methods for additional analyses, such as subgroup analyses and adjusted analyses
No corresponding text identified

13a For each group, the numbers of participants who were randomly assigned received intended treatment and were analyzed for the primary outcome
No corresponding text identified

13b For each group, losses and exclusions after randomization, together with reasons
No corresponding text identified

14a Dates defining the periods of recruitment and follow-up
No corresponding text identified

14b Why the trial ended or was stopped
No corresponding text identified

15 A table showing the baseline demographic and clinical characteristics for each group
No corresponding text identified

17b For binary outcomes, presentation of both absolute and relative effect sizes is recommended

No corresponding text identified

18 Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory

No corresponding text identified

19 All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)

No corresponding text identified

21 Generalizability (external validity, applicability) of the trial findings

No corresponding text identified

23 Registration number and name of trial registry

No corresponding text identified

24 Where the full trial protocol can be accessed, if available

No corresponding text identified

25 Sources of funding and other support (such as the supply of drugs), the role of funders

No corresponding text identified

2. Some items appear to be included in your manuscript

Your manuscript may include the following items recommended the [SPIRIT / CONSORT] guidelines. Nonetheless, we suggest you review the selected text below to confirm that your manuscript includes all components in each [SPIRIT / CONSORT] item.

16 For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups

This process results in about 684K training triples and 145K validation triples.

17a For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)

Table 1 presents the main results corresponding to our evaluation tasks (described in Â§3).

Overall, we observe substantial improvements across all tasks with average performance of 80.0 across all metrics on all tasks which is a 3.1 point absolute improvement over the next-best baseline.

Our evaluation of the learned representations on predicting user activity is shown in the "User activity" columns of Table 1.

SPECTER achieves a MAP score of 83.8 on the co-view task, and 84.5 on coread, improving over the best baseline (Citeomatic in this case) by 2.7 and 4.0 points, respectively.

20 Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, the multiplicity of analyses

14 Venue information in our data came directly from publisher provided metadata and thus was not normalized.

22 Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence

This is because for this task the embeddings are used along with several other informative features in the ranking model (described under task-specific models in Â§4), meaning that embedding variants have less opportunity for impact on overall performance.

In a dataset of 4,113 clicks, we found that SPECTER ranker improved clickthrough rate over the baseline by 46.5%,

demonstrating its superiority.

We emphasize that our citation-based pretraining objective is critical for the performance of SPECTER; removing this and using a vanilla SciB-ERT results in decreased performance on all tasks.

One possible explanation is that author names are sparse in the corpus, making it difficult for the model to infer document-level relatedness from them.

While there could be other potential ways to include hard negatives in the model, our simple approach of including citations of citations is effective.

As illustrated in Table 3, without any additional final task-specific fine-tuning, SPECTER still outperforms a SciBERT model fine-tuned on the end tasks as well as their multitask combination, further demonstrating the effectiveness and versatility of SPECTER embeddings. 15