**Paper 2**

**Summary**

The paper describes a methodology for extracting relevant evidence and categorizing claims into predetermined classes using Long Short-Term Memory (LSTM) neural networks, which are known for their aptitude in handling sequential data and have achieved success in various NLP tasks. The paper employs pre-trained word embeddings from GloVe to encode the input data, and explores the utilization of LSTM for both evidence retrieval and claim classification, discussing the LSTM architecture, dataset, preprocessing measures, and results. The results demonstrate satisfactory performance in both tasks. Overall, the paper offers a promising solution for evidence retrieval and claim classification, applicable to information retrieval and fact-checking domains.

**Strengths**

The paper is well-written and organized for the previous parts, allowing readers to follow the ideas and methodology easily. It is a pleasure to read this report, full of professional, concise, and accurate adjectives conveying the content concisely and accurately.

Furthermore, the paper gives considerable insight into the dataset, especially a thorough investigation of the evidence data. And the description of the preprocess is clear and compact, showing a deep understanding of the techniques of the author.

Besides, it is admirable that the great physical and financial efforts the author has put into this project, and the exciting outcomes once again prove that these efforts are worthwhile.

Besides, the report showcases a sense of critical thinking. The assumption “the model is trained to retrieve all relevant evidence documents given a claim” (or the assumption of perfect performance) is of great importance but seldom do people point this out.

**Weaknesses**

While the writing clarity and professional wording are commendable, possible improvements include the re-organization of the passage structure, more experiment results to show the comparison of different methods, and more effective use of tables and figures.

Firstly the structure of this paper can be adjusted to a more complete and organized one. Specifically, the abstract is too long, and a conclusion summarizing the outcomes above plus more information about future improvements should be added. Also, the paper uses a great passage to justify the usage of the GLOVE pre-trained model, which is full of logic but somewhat lengthy for this report. It would be better to break one huge paragraph into smaller ones for readers to read (such as suction 4.1 and 4.2) though the discourses of the content are identifiable through adverb words. Adding more subtitles for each model within larger subtitles would help the report to be clearer.

Secondly, more details related to the experiment results should be added. For example, it mentioned whether removing stop words should depend on different situations, but it is still confusing what method was used. More details can be provided on the difference between hyperparameter "lr" and the way the Adam optimizer changes the learning rate since the paper has mentioned the optimizer's function specifically. And the report has not talked about how to retrieve documents that are related to the claim, only the binary classification part was mentioned. What’s more, lemmatization is introduced as a technique to improve accuracy, but no experimental or quantified results were given to justify the claim.

Lastly, tables and figures can be employed to present important information, making the results and key details interpretable. For example, a multicolumn table rather than a simple table with values squeezed together (Table 1) can be utilized to show the

results of parameter tuning more nicely and neatly. And the configuration with the best score can be highlighted to emphasize the final results. Also sometimes it would be better to illustrate the architecture of a network using graphs rather than prolix while informative words.