Skoltech
Skolkovo Institute of Science and Technology

Jury Member Report – Doctor of Philosophy thesis.

Name of Candidate: Aleksandr Rubashevskii

PhD Program: Computational and Data Science and Engineering

Title of Thesis: Efficient active learning and fact-checking via predictive uncertainty quantification for natural language processing

Supervisor: Assistant Professor Maxim Panov

Name of the Reviewer:

I confirm the absence of any conflict of interest Ivan Oseledets	**Date: 01-12-2024**

The purpose of this report is to obtain an independent review from the members of PhD defense Jury before the thesis defense. The members of PhD defense Jury are asked to submit signed copy of the report at least 30 days prior the thesis defense. The Reviewers are asked to bring a copy of the completed report to the thesis defense and to discuss the contents of each report with each other before the thesis defense. *If the reviewers have any queries about the thesis which they wish to raise in advance, please contact the Chair of the Jury.*
Reviewer's Report

Summary of Main Results:

The thesis is devoted to the uncertainty quantification (UQ) for. Na

tasks. The first major contribution is the development of a new batch active learning algorithm called Large BatchBALD (LBB). LBB provides an efficient approximation of the BatchBALD method, allowing for the selection of diverse batches of samples for annotation while significantly reducing the computational complexity compared to the original BatchBALD. The author demonstrates the effectiveness of LBB on both text and image classification tasks.

The second main contribution is a novel token-level uncertainty quantification method called Claim Conditioned Probability (CCP). CCP is designed for fact-checking the outputs of large language models, as it is able to distinguish between different types of uncertainty (e.g. uncertainty about the claim type vs. uncertainty about the specific claim value) and focus only on the uncertainty relevant for detecting factual errors. CCP is shown to outperform several baselines on the task of hallucination detection in language model-generated biographies.

Finally, the thesis presents the construction of Factcheck-Bench, a comprehensive benchmark for evaluating end-to-end fact-checking systems on language model outputs. The benchmark includes annotations at multiple levels of granularity (claim, sentence, document) and covers both detection and correction of factual errors.

Novelty and Significance:

The methods and benchmark developed in this thesis represent significant advancements in their respective areas. The LBB active learning algorithm provides an important practical improvement over existing batch-based approaches. The CCP uncertainty quantification technique offers a novel perspective on separating relevant from irrelevant uncertainty for specific NLP tasks. And the Factcheck-Bench benchmark fills an important gap in the evaluation of fact-checking systems, going beyond previous work that has focused more narrowly on specific subtasks.

Overall Structure:

The thesis is reasonably structured, but it can be improved a little bit (see the comments).

Comments:

1) It is not clear, what is the personal contribution of the author of the thesis to the joint work. This could be highlighted in more details.
2) Some of the conclusions come directly from the papers, for example (page 52, section 3.3: 'In this work, (not chapter)
3) Chapter 4: The term NLI is not defined in the text.
4) The models used in the chapters (like Vicuna-13B) are quite outdated. What will happen if they are replaced by more modern models?
5) As a ground truth for FactCheck benchmark, most of the work is done by GPT-4 model, as far as I can understand. How we can trust the output of this model in benchmark construction? Is there any 'gold standard' that has been cross-checked by other sources?

Overall, I think this is an interesting and solid work for a PhD thesis.

Provisional Recommendation
I recommend that the candidate should defend the thesis by means of a formal thesis defense
I recommend that the candidate should defend the thesis by means of a formal thesis defense only after appropriate changes would be introduced in candidate's thesis according to the recommendations of the present report
The thesis is not acceptable and I recommend that the candidate be exempt from the formal thesis defense