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: Dr. Mikhail Burtsev

I confirm the absence of any conflict of interest 	**Date: 19-11-2024**

Reviewer's Report

The thesis is well written and has clear structure. This thesis investigates the important problem of reliability and constraints in the performance of language models, which is a bottleneck when using NLP models for a wide range of applications. The author addresses this problem using uncertainty quantification, which is highly relevant for the problem under consideration.

The introduction and related work effectively outline the importance of uncertainty estimation in NLP and the challenges associated with it, providing a strong foundation for the subsequent chapters. The main chapters are well-structured, the author gives a mathematical derivation of methods with intuitions behind it. The methods given are compared experimentally with modern alternatives. The experimental results validate the effectiveness of the proposed methods, demonstrating improvements over existing techniques. The practical applications of the results are evident.

The author's main contribution for uncertainty estimation is presented on two tasks: active learning and fact-checking. In the case of active learning, an approximation of the BatchBALD is presented in order to speed up the algorithm computation. In the case of fact-checking, a new method of uncertainty estimation for language models called Claim Conditioned Probability (CCP) is presented, in which only the uncertainty about the generated fact is considered and uncertainty about the stylistic features of the language is not taken into account. Also, for fact-checking, the author presents the annotation methodology and the corresponding processed dataset for the ChatGPT model.

The dissertation is supported by high-quality publications, indicating the peer-reviewed validation of the research. The list of publications includes such conferences as SIAM SDM, ACL and EMNLP. Unfortunately, there is no information provided in the thesis about specific contributions of the author in the above-mentioned publications, the majority of which have many authors. I assume that the results presented in the dissertation are produced by the author.

Specific comments regarding the text of the dissertation are provided below.

1. On the p. 45 in the caption of Fig. 3-1 the author states that "LBB clearly outperform other competitors". I was not able to see any substantial evidence for this claim. LBB is aligned with Random sampling. In the overall, results on the Fig. 3-1 are hard to interpret due to the lack of details. Is color shades represent confidence intervals? How were these confidence intervals calculated? Are there any statistically significant differences between data points? Until these questions are answered, it is hard to make clear sense of the results.
2. On the p. 46 the Table 3.2. and many following tables show confidence intervals. I was not able to find in the text how these confidence intervals were calculated.
3. For active learning in image domain models were retrained from scratch after addition of new samples. Is it common approach to active learning in the image domain? Will conclusions from evaluation be different if models fine-tuned but not re-trained after addition of new samples?
4. P. 72, Table 4.12. It would be good to know how different parts of the full pipeline influence a runtime. I suggest including values for CCP with claim extraction and matching if possible.

In summary, the dissertation is a high-quality research work, with significant scientific, methodological, and practical relevance.

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*