Improving the transformer-based ChatBERT

At NLB, we are committed to protecting the environment. To that end, we publish annual ESG (Environmental, Social, and Governance) reports that detail the sustainable development efforts undertaken by our employees. At NLB DigIT, we have developed a transformer-based question-answering (QA) model that allows users to ask questions and obtain answers related to the reports' content. We have created a dataset of questions, answers, and contexts from which the answers were manually retrieved or extracted using existing QA tools. We have fine-tuned a BERT model on the dataset and achieved xx% accuracy on the test set. Can you suggest ways in which we can further improve the model?

Your task is to try to improve the existing question answering system trained on the questions and answers extracted from the bank’s annual ESG sustainability reports, available online (and in the local folder “reports”):

* <https://www.nlb.si/sustainability_report_2022.pdf>
* <https://www.nlb.si/sustainability-report-2020.pdf>

In the folder “dataset” there are two Excel files:

* questions\_answers\_train.xlsx, containing 544 questions and answers
* questions\_answers\_test.xlsx, containing 52 questions and answers

Questions, answers and contexts where extracted using the HayStack platform:

<https://haystack.deepset.ai/overview/demo>

The code for the QA system was written in Python and is available in the Jupyter notebook which can be found in the local directory. Your task is to improve the performance of the existing model, having the evaluation code at the bottom of the notebook.

### Ideas:

* clean data,
* experiment with parameters,
* try different models and architectures,
* anything else that comes up to your mind!