Challenge: LLM for Insurance Claim Automation

Scenario: You work for an InsurTech startup which is looking to develop a solution leveraging LLMs for Insurance companies that receive numerous insurance claims every day in various formats such as the incident description, communication history, and other relevant notes. The Insurance Company (Initial Target Customer) wants to use LLMs to automate the initial claim processing by categorising the claims into different categories (e.g., Auto, Home, Life, Health) and extracting key claims payout details from the claims to process a claim payout.

Part 1:

- Explain briefly how a transformer-based model can work in this use case, the benefits and why it might be suitable for this scenario
- Discuss the potential challenges of using LLMs in the insurance sector
- How would you approach designing and implementing an LLM-based solution in a sector like insurance where the accuracy and reliability of results are critical
- Given the model will be used in a regulated environment, elaborate on ways to ensure that the models comply with privacy, security, and regulatory guidelines

Part 2:

- Use a pre-trained LLM
- Preprocess a subset of synthetic claims data and convert it into a format suitable for training the LLM
    - You must generate the data you will use to demonstrate the system by looking at examples from the internet. Add the files used to the repository.
- Implement appropriate tests to ensure the accuracy of the model and its ability to consistently extract correct information from the text
- Evaluate the performance of your model. Discuss the evaluation metrics used and their interpretation
- Ensure your model, data handling, and results comply with privacy, security, and regulatory guidelines relevant to the insurance sector

You will need to submit:

All source code developed for the task should be published to GitHub

A short report (2-5 pages) summarising your findings, including figures where necessary.

This should include a description of the model architecture, preprocessing steps, performance metrics, and a discussion on models privacy, compliance and reliability measures

Note: The challenge is designed to assess your understanding of LLMs and your ability to apply them to a practical problem within the insurance sector. Your methodology, ability to discuss and justify your choices, and consideration of industry-specific constraints are just as important as your final solution.

Please be prepared to discuss your solution in a follow-up interview, where you might be asked to justify your choices, propose improvements, or discuss alternative solutions.