

Project Proposal: Question Answering — Task-Specific Models vs. LLMs

Course Project

Code of Honor. All external resources used in the project, including research papers, open-source repositories, datasets, and any content or code generated using AI tools, e.g., ChatGPT, GitHub Copilot, Claude, Gemini, must be *clearly cited* in the final submission. The final report must also include *a clear breakdown of individual group member contributions*. Any lack of transparency in the use of external resources or in reporting group contributions will be considered academic dishonesty and will significantly impact the final evaluation.

Topic	Question Answering — Task-Specific Models vs. LLMs
Dataset	Recipe-MPR multi-perspective recipe QA

BACKGROUND Large language models (LLMs) perform strongly on open-domain question answering, yet their general-purpose nature can miss task-specific nuances. In specialized domains, fine-tuning a focused model often yields gains in efficiency, interpretability, and reliability. The Recipe-MPR dataset, containing 500 user queries with five answer variations apiece, offers a benchmark for multi-perspective responses in the recipe domain and enables a grounded comparison between specialized and general-purpose approaches.

OBJECTIVE Fine-tune a task-specific transformer model on Recipe-MPR to surpass the 65% baseline accuracy while benchmarking performance, robustness, and efficiency against prompted LLM outputs.

MOTIVATION By contrasting a domain-adapted model with a general-purpose LLM, we aim to surface the trade-offs between specialization and broad capability. Understanding when a dedicated model outperforms or complements a large foundation model can inform deployment decisions in domains demanding efficiency, transparency, or domain fidelity.

REQUIREMENTS The final submission should address the following requirements while the details can be freely decided by the group members.

- Model Training:** Fine-tune a pretrained transformer (e.g., BERT, RoBERTa, DistilBERT) on Recipe-MPR to capture recipe-domain subtleties.
- Evaluation:** Report accuracy and F1-score on a held-out test split to measure task-specific model performance.
- Comparison:** Prompt a state-of-the-art LLM with identical queries, score its responses using the same metrics, and compare with the fine-tuned model.
- Analysis:** Provide qualitative and quantitative analyses to highlight scenarios where the specialized model excels or complements the LLM baseline.

MILESTONES The following milestones are to be accomplished through semester.

1. **Weeks 1–2:** Conduct literature review on recipe QA, finalize dataset handling, and prepare the development environment.
2. **Weeks 3–4:** Fine-tune the selected transformer model variants on Recipe-MPR.
3. **Week 5:** Evaluate the task-specific model and benchmark against LLM-generated answers.
4. **Week 6:** Perform comparative analysis and draft key findings.
5. **Final Week:** Complete the written report and prepare the project submission materials.

EXPECTED OUTCOMES We anticipate the fine-tuned model achieving accuracy comparable to or exceeding that of prompted LLMs within the recipe domain. The project should demonstrate the efficiency, domain adaptability, and interpretability gains of specialized models while clarifying the circumstances in which reliance on general-purpose LLMs remains advantageous.

SUBMISSION GUIDELINES The main body of work is submitted through Git. In addition, each group submits a final paper and gives a presentation. In this respect, please follow these steps.

- Each group must maintain a Git repository, e.g., GitHub or GitLab, for the project. By the time of final submission, the repository should have
 - Well-documented codebase
 - Clear README.md with setup and usage instructions
 - A requirements.txt file listing all required packages or an environment.yaml file with a reproducible environment setup
 - Demo script or notebook showing sample input-output
 - *If applicable*, a /doc folder with extended documentation
- A final report (maximum 5 pages) must be submitted in a PDF format. The report should be written in the provided formal style, including an abstract, introduction, method, experiments, results, and conclusion.
Important: Submissions that do not use template are considered *incomplete*.
- A 5-minute presentation (maximum 5 slides including the title slide) is given on the internal seminar on Week 15, i.e., Dec 8 to Dec 12, by the group. For presentation, any template can be used.

FINAL NOTES While planning for the milestones please consider the following points.

1. You are encouraged to explore innovative approaches to conditioning or generation as long as the core objectives are met.
2. While computational resources are limited, carefully chosen datasets and training setups can make even diffusion models feasible. Trade-offs, e.g., resolution, training steps, are expected and should be justified.
3. Teams are expected to manage their computing needs and are advised to perform early tests to estimate runtime and training feasibility. As graduate students, team members can use facilities provided by the university, e.g., ECE Facility. Teams are expected to inform themselves about the limitations of the available computing resources and design the model accordingly.

REFERENCES