Data Science Report: Course Suggestor Agent

1 Fine-Tuning Setup

This report details the fine-tuning process for the **QueryParsingAgent**, a mandatory component of this project [**?**].

1.1 Model Selection

Model: `microsoft/Phi-3-mini-4k-instruct`

Reasoning: This model was chosen as it offers a state-of-the-art balance of high performance and low computational cost. Its 3.8B parameters are small enough to be fine-tuned efficiently on a single consumer GPU (like a T4 in Google Colab) using 4-bit quantization (QLoRA).

1.2 Fine-Tuning Target & Rationale

Target: The goal was Task Specialization. A base LLM is not reliable for generating structured JSON that conforms to a specific, rigid database schema.

Rationale: By fine-tuning the model, we teach it to act as a "Natural Language to JSON Query" translator. This is more reliable than complex prompt engineering. The model learns the specific mappings between user intent (e.g., "easy course") and our database logic (`{"sortBy": "AGP", "sortOrder": "desc"}`).

1.3 Data

- **Training Data:** A custom dataset of 104 prompt-completion pairs was created manually (see `finetune_data.jsonl`). Each entry maps a sample user question to its ground-truth JSON query.
- **Knowledge Base:** The agent's knowledge comes from `final_database.json`, a file created by merging the `Course_Schedule_2024-25-2.pdf` with historical grading data from four other official DOAA reports [**?**, **?**, **?**, **?**, **?**].

1.4 Method

Technique: QLoRA (4-bit Quantization with LoRA)

Hyperparameters:

- r: 16
- `lora_alpha`: 32

- learning_rate: 2e-4
- num_train_epochs: 3
- optimizer: paged_adamw_8bit

2 Evaluation Methodology and Outcomes

Evaluation was performed on both the fine-tuned component (Planner) and the end-to-end agent [**?**].

2.1 Quantitative Outcome (Planner Accuracy)

Methodology: An Exact Match Accuracy (EMA) test was conducted. A new test set of 20 unseen prompts was run through the fine-tuned `get_json_query` function. The generated JSON string was compared to the expected ground-truth string.

Results:

[PASTE YOUR ACCURACY SCORE HERE, e.g., 18/20 = 90.0%]

(See `quantitative_evaluation_log.txt` for detailed results.)

2.2 Qualitative Outcome (Agent Reliability)

Methodology: Five complex, multi-constraint queries were run through the full `run_course_suggester` to assess its ability to satisfy constraints and provide a useful final answer.

Results (Chat Logs):

(Paste the output from your "Qualitative Evaluation" cell here. This serves as your qualitative log.)

```
[EXAMPLE LOG]
--- User Request ---
I need an easy morning class about 'Structures'.

--- Agent Response ---
Based on your request for an easy morning class about 'Structures',
I found one course:

STRUCTURAL ANALYSIS (CE272): This course has 'Average Difficulty' and
meets on Monday (11:00-12:00) and Tuesday/Friday (12:00-13:00).
(...etc...)
```

References