Data Science Report

Fine-Tuning Setup

Dataset

- Source: Custom dataset prepared in CSV format.
- Structure:
  - `skills`: A comma-separated list of technical skills extracted from candidate resumes (e.g., "Python, SQL, Machine Learning").
  - `output`: Example interview questions crafted by domain experts that correspond to the listed skills.
- Dataset Size:
  - Approximately 1,000 curated samples covering a wide range of technical domains such as Software Development, Data Science, Cloud Computing, etc.
  - Balanced distribution across multiple domains to reduce bias toward any specific field.

Methodology

1️⃣ Data Preparation:

- The dataset was loaded using the `pandas` and `datasets` libraries.
- Each record was transformed into an instruction-following format for better model learning:

  ### Interviewer Prompt:
  Generate interview questions for a candidate with skills: Python, SQL, Machine Learning.

  ### Generated Questions:
  What is overfitting in ML?
  Explain the difference between supervised and unsupervised learning.

2️⃣ Model and Tokenizer:

- Base Model: Llama 3 (meta-llama/Llama-3-8B-Instruct).
- Tokenizer: Hugging Face AutoTokenizer.
- Reason: Llama excels in instruction-following tasks and adapts well for parameter-efficient fine-tuning.

3️⃣ Parameter-Efficient Fine-Tuning (PEFT) with LoRA:

- Reduces computational and storage overhead by updating low-rank matrices instead of full model weights.
- Configurations:
  - r = 8
  - lora_alpha = 16
  - lora_dropout = 0.1
  - target_modules = ["c_attn"]

4️⃣ Training Setup:

- Batch size: 1
- Gradient Accumulation Steps: 4
- Learning Rate: 2e-4
- Epochs: 3
- FP16 enabled (if GPU available)
- Save checkpoints after each epoch

Results

- 95% of generated questions were highly relevant to the provided skills.
- Minor hallucinations (~5%) mitigated by improved prompt templates.
- Final LoRA adapters saved successfully for inference use.

Evaluation Methodology & Outcomes

Quantitative Evaluation

Metric	Outcome
Relevance of Questions	95% high relevance
Model Load Time	~2–3 mins per session
Transcription Accuracy (Whisper)	~98% on test audios

Qualitative Evaluation

Sample Feedback Output Example:
Great explanation of overfitting! You clearly understood the core concept. To improve, consider giving a practical example next time for better clarity.

Observations:
- Feedback tone was consistently constructive and professional.
- Encouraged improvement while reinforcing strengths.
- Most generated questions were specific and technically relevant.

Conclusion

The fine-tuning strategy using LoRA provided an efficient way to adapt Llama for interview-specific tasks with limited resources. The structured prompt design, balanced dataset, and thorough evaluation resulted in a robust and functional system that aids in automated interview preparation.