LLM Overview

□□□□□□□□□□ □□□□□□□□□ □□□□□□□-□□□□□□□□□□□:

While fine-tuning adapts LLMs to specific tasks, its limitations are becoming clear. High computational costs, potential "catastrophic forgetting," and challenges achieving deep domain expertise call for innovative approaches.

□□□□□□□□□ □□□□□□□□□□□□□□□□□□-□□□□□□□□□□□□□□□□□ □□□□□□□□□□□□□□□□□□□□□ (□□□□□□):
RAG equips LLMs with an open book of relevant information, retrieving key passages from a knowledge base to provide factual context and enhance responses beyond the LLM's training data. This leads to:

- Improved Accuracy: Reduces hallucinations and factual errors by grounding responses in real-world information.
- Domain Expertise: Injects domain-specific knowledge for richer, more targeted outputs.
- Reduced Training Costs: Focuses fine-tuning on generation, requiring less labeled data.

□□□□□□□□□□□□□□□□□□□□□□□□□□□□ □□□□□□□□□□□ □□□□□□□□□□□□□□□□□□ (□□□□□□):
While RAG excels at accessing external knowledge, sometimes we need LLMs to become true subject-matter experts. ITS leverages domain-specific data and instructions to:

- Tailor LLM Responses: Focuses on specific tasks and knowledge within your chosen domain.
- Enhance Task-Specific Accuracy: Optimizes the LLM for tasks like question answering or summarization.
- Explainable Results: Shows how the LLM reached its conclusions, improving interpretability.

□□□□□□□□□□□ □□□□□□, □□□□□□, □□□□□□ □□□□□□□□□□□□□□□□:
The LLM landscape continues to evolve with noteworthy methods like:

- Parameter-Efficient Fine-Tuning (PEFT): Reduces costs and mitigates "catastrophic forgetting" by adapting only select parameters per task.
- Dense/Sparse Passage Retrieval: Retrieves information at different granularities for summarization or entity identification.
- Prompt Engineering & Templates: Craft prompts to guide the LLM or use templates for consistent output formats.

□□□□□□□□□□□□□□□□ □□□□□□ □□□□□□□□□□ □□□□□□□□□□□□□:
Consider the nature of your task, resources available, desired accuracy, and domain expertise needed. Experimentation and strategic combinations of methods can unlock the full potential of LLMs for your projects.

Revision #2
Created 23 April 2024 11:43:47 by sedawk

Updated 10 February 2025 16:35:32 by sedawk