Magnimind Academy LLM Workshop
Capstone Project Guideline

Objective

The capstone project is an opportunity for students to apply the knowledge gained from the six-week workshop on Large Language Models (LLMs). Students will design, implement, and present a project that demonstrates their understanding of one or more topics: LLM fundamentals, training/fine-tuning, semantic search, Retrieval-Augmented Generation (RAG), prompt engineering, or agentic systems. In Week 6, students will work through a sample project end-to-end with instructors, and in Week 7, they will present their own projects.

Project Requirements

Scope

The project must incorporate at least one key concept from the workshop (e.g., fine-tuning an LLM, building a RAG system, designing an agent with tools, or advanced prompt engineering). Projects should aim to solve a practical problem or demonstrate a creative application of LLMs.

Deliverables

- **Implementation**: A working prototype or proof-of-concept (e.g., code, model, or application) showcasing the chosen LLM application.

- **Documentation**: A 2–3 page report (PDF) detailing:

    - Problem statement and project goals.

    - Methodology, including tools, datasets, or models used.

   – Challenges faced and how they were addressed.

   – Results and evaluation (qualitative or quantitative).

   – Future improvements or extensions.

- **Presentation**: A 10-minute slide presentation (8–12 slides) to be delivered during the Week 7 Showcase, covering:

   – Introduction and motivation.

   – Technical approach and implementation.

   – Results or demo.

   – Lessons learned and future work.

Technical Constraints

- Use open-source tools, libraries, or models (e.g., Hugging Face, LangChain, Llama) where possible.

- Ensure the project is executable on standard hardware (e.g., a laptop with 16GB RAM or cloud-based solutions like Google Colab).

Project Ideas (Optional)

Students may choose their own project or select from the following:

- **Semantic Search System**: Build a search engine for a specific domain (e.g., academic papers) using embeddings and vector databases.

- **RAG Application**: Create a question-answering system that retrieves relevant documents and generates answers using an LLM.

- **Agentic System**: Design an agent that uses tools (e.g., calculator, web scraper) for tasks like travel planning or data analysis.

- **Fine-Tuned Model**: Fine-tune a small LLM on a niche dataset (e.g., customer support chats) for tasks like sentiment analysis.

- **Prompt Engineering Showcase**: Develop optimized prompts for a creative task (e.g., story generation) and compare performance across models.

Timeline

- **Week 5**: Submit a one-paragraph project proposal (due August 20, Wednesday 5pm PST) that outlines the problem, the approach, and the expected outcomes. Share your project proposal with instructors via Slack. Feedback will be provided within 48 hours.

- **Week 6**: Participate in an end-to-end project walkthrough with instructors to guide your project development. Begin implementing your project.

- **Week 7**: Complete the project implementation, documentation, and presentation slides. Submit deliverables by the morning of the Week 7 Showcase.

Evaluation Criteria

Projects will be evaluated based on:

- **Technical Depth (30%)**: Demonstrates understanding of LLM concepts and effective use of workshop techniques.

- **Creativity and Impact (25%)**: Addresses a meaningful problem or showcases an innovative application.

- **Implementation (25%)**: Produces a functional prototype with clear results or a working demo.

- **Presentation and Documentation (20%)**: Clearly communicates the project's purpose, approach, and outcomes in both written and oral formats.

Week 6 Project Walkthrough

- **Purpose**: In Week 6, instructors will guide students through a sample LLM project from start to finish, covering planning, implementation, and evaluation. This will serve as a model for students' own projects.

- **Activities**: Students will follow along, ask questions, and begin applying the demonstrated techniques to their own projects. Use this session to refine your project plan and troubleshoot initial challenges.

- **Preparation**: Review your Week 5 proposal and bring specific questions or ideas to the walkthrough.

Week 7 Showcase Format

- **Date**: Week 7, during the scheduled workshop session.

- **Format**: Each student (or group, if collaborative) will present for 10 minutes, followed by 5 minutes of Q&A.

- **Submission**: Upload slides (PDF) and documentation (PDF) to the workshop platform before the session. Code or prototypes should be accessible via a GitHub link or similar.

- **Demo**: If applicable, include a live or recorded demo in the presentation. Test demos thoroughly to ensure they run smoothly.

Tips for Success

- Use the Week 6 walkthrough to clarify technical approaches and avoid common pitfalls.

- Start implementation early in Week 6 to allow time for debugging and refinement.

- Keep slides concise and visually clear, using diagrams or screenshots to illustrate key points.

- Practice your presentation to stay within the 10-minute limit.

- If stuck, revisit workshop materials or consult instructors during Week 6.

Resources

- **Tools**: Hugging Face Transformers, LangChain, Google Colab, or local Python environments.

- **Datasets**: Public datasets from Kaggle, Hugging Face Datasets, or domain-specific sources (e.g., arXiv for research papers).

- **Documentation**: Refer to workshop notes for code snippets, model configurations, or prompt templates.