Proposal: EE 517 Course Revision

Banghua Zhu, Maryam Fazel

Context

Previous course information:

- Title: Continuous-Space Language Processing
- Description: Introduction to human language technology, with in-depth coverage of **continuous-space statistical models** of language and application to natural language processing tasks. Methods covered include low rank distributional representations, neural networks, and log bilinear statistical models, which are leveraged for language modeling, similarity scoring, classification, and translation/generation. Prerequisite: E E 505.

New course description:

- Title: Introduction to Large Language Models
- Description: EE 517: Introduction to Large Language Models is an introduction to the principles and applications of large language models (LLMs). Topics include the **architecture, model training and fine-tuning, evaluation, efficient serving, ethical considerations, and real-world applications**.

Topics

- Introduction to language models and self-supervised learning.
- The transformer architecture and self-attention mechanism.
- Pre-training, fine-tuning, and prompt engineering.
- Evaluation metrics and benchmarking for LLMs.
- Scaling laws and emergent properties of large models.
- Efficiency techniques (quantization, distillation, retrieval-augmented generation).
- Agentic Applications: agent framework (DSPy, MCP, Langchain) and compound AI systems
- Ethical considerations: bias, fairness, misinformation, and environmental impact.
- LLM deployment and real-world considerations.
- Future directions and open challenges in LLM research.

Justification: The Need for New LLM Course

Modernization of Curriculum: The new course content reflects the rapid evolution of language modeling technology from **traditional statistical models** to **transformer (and other new architecture)- based** large language models, aligning with current state-of-the-art methods in both research and industry.

Broader and Deeper Scope: While the previous course emphasized continuous-space models and specific neural architectures, the new course provides a comprehensive view of LLMs, including architecture, pre-training and post-training techniques, evaluation, agentic applications, and safety issues.

Alignment with Industry Needs: LLMs are now foundational to many AI products and services. This course prepares students with hands-on skills and theoretical knowledge that are directly relevant to roles in tech companies, startups, and research labs.

[1]The MCD for 445 is under revision to reflect these changes as well.