Course Title: Foundations of Large Language Models
Instructor: Samet Oymak

Course description: This course aims to provide students with an in-depth understanding of state-of-the-art language models, focusing on transformers and models like ChatGPT. Through comprehensive technical study, students will delve into algorithmic, systems-level, and theoretical aspects of these models. Beyond the technical foundation, the course will also explore vital societal considerations, from ensuring models align with human values to examining AI safety measures. By the end of the course, students will not only grasp the intricacies of modern language models but also appreciate their broader implications and challenges in real-world applications.

Course syllabus: At a high-level, the course will cover three main areas on large language models (LLM):
(1) Theoretical foundations of transformers and LLMs
(2) Systems and Algorithms for LLMs
(3) Safety and Trustworthiness of LLMs.
Systems and Algorithms, and Societal aspects of large language models.

Detailed weekly schedule:

Week 1: Introduction to Language Models and Transformers
- Class 1: Overview of NLP and Historical Development of Language Models
- Class 2: Introduction to Transformer Architecture and Evolution from RNNs

Week 2: Fundamentals of Transformers
- Class 1: Self-attention Mechanism and Positional Encoding
- Class 2: Multi-head Attention, Feed-forward Networks, and Residual Connections

Week 3: Advanced Transformer Variants and ChatGPT
- Class 1: Deep Dive into GPT, BERT, and T5
- Class 2: ChatGPT: Architecture, Training, and In-context Learning

Week 4: Scaling and Efficiency in Transformers
- Class 1: Scaling Issues: Model Size, Computation, and Memory
- Class 2: Model Compression, Efficient Deployment, and TF-specific Optimization

Week 5: Theoretical Foundations and Approximations
- Class 1: Understanding Transformers' Capacity and Approximation Ability
- Class 2: Theory of Attention and Attention Approximations

Week 6: Systems-level and Theoretical Considerations
- Class 1: Systems-level Optimizations for Training Large Models
- Class 2: Generalization in Transformers and Theoretical Bounds

Week 7: Compositional and Sequential Learning

- Class 1: Emergent Abilities and Scaling Laws in Transformers
- Class 2: Compositional and Sequential Learning Techniques with Transformers

Week 8: Prompting Techniques and Theoretical Insights

- Class 1: Theoretical Properties of Prompting Techniques
- Class 2: Transfer Learning, Fine-tuning, and Applications in NLP

Week 9: Efficient Transformer Architectures

- Class 1: Non-attention architectures
- Class 2: Low-rank and sparse approximation of self-attention

Week 10: Instruction Tuning and Alignment

- Class 1: Alignment to Human Values and Instruction tuning
- Class 2: Reinforcement Learning from Human Feedback

Week 11: Language models as agents

- Class 1: Language models as tool-users
- Class 2: Multiagent LLMs and Human-AI Interaction

Week 12: Societal Implications

- Class 1: Transparency, Explainability, Watermarking
- Class 2: Economical and Societal Shifts Due to Language Models

Week 13: Advanced AI Safety and Robustness

- Class 1: Bias, Fairness, and Representation in Language Models
- Class 2: Hallucination, Risk in Decision Making, Concerns on Existential Risk

Week 14: Final Projects and Presentations

- Class 1: Student Presentations on Course Projects
- Class 2: Student Presentations on Course Projects

Grading:

- **Class presentation and report:** 30%
- **Project presentation and report:** 30%
- **Midterm:** 20%
- **Participation:** 10%
- **Project feedback:** 10%