

Large Language Model

1. Lecture Hours: 32
2. Laboratory Hours: 0
3. Colloquial Hours: 0
4. Individual Study Hours: 32
5. Credits: 2
6. Occurrence: SummerCourse
7. Prerequisites: Programming-related courses*, Machine Learning
8. Course Description:

This course explains the fundamental principles and practical workflows of large language models. Centered on the Transformer architecture, the course examines how modern large language models learn, generate, and align natural language through large-scale pre-training and post-training refinement. The course also introduces efficient training, parameter-efficient fine-tuning, quantization, and inference acceleration methods for building and adapting large language models in practical scenarios. Upon completing this course, students should be able to understand the theoretical foundations of large language models and apply appropriate techniques for model training, adaptation, and deployment.

9. Course Outcomes:

By the end of successful completion of this course, the student will be able to:

- (1) Explain the fundamental principles and Transformer architecture of large language models.
- (2) Describe the training process of large language models, including tokenization, pre-training, optimization, and scaling.
- (3) Analyze efficient training and distributed computing methods for large language models.
- (4) Apply alignment and fine-tuning techniques to adapt large language models to different tasks.
- (5) Use inference optimization methods to support practical deployment of large language models.

10. Course Content, Laboratories and Laboratory Hours, Colloquial Hours

- (1) The Transformer Architecture and Model Variants (6 Class Hours)
 - Classroom 6 hours
- (2) Pre-training Dynamics and Scaling Laws (4 Class Hours)
 - Classroom 4 hours
- (3) Efficient Training and Distributed Systems (6 Class Hours)
 - Classroom 6 hours
- (4) Alignment and Post-Training Methods (6 Class Hours)
 - Classroom 6 hours
- (5) Parameter-Efficient Fine-Tuning and Model Adaptation (4 Class Hours)

- Classroom 4 hours
- (6) Inference Optimization and Deployment (6 Class Hours)
- Classroom 6 hours

11. Grading:

Assignment: 30%

Course Report: 25%

Final Exam: 45%

12. Text & Reference Book:

Raschka, S. Build a Large Language Model (From Scratch). 1st edition. Manning Publications, 2024. ISBN: 9781633437166.

13. Course Lecturer: YiDa Xu, Kan Li

Course Lecturer: Kan Li