4190.773: Topics in Artificial Intelligence

Large Foundational Models and Physical World Models for Multimodal AI

Sangdoo Yun[1,2] Jin-Hwa Kim[1,2]
[1]NAVER AI Lab [2]Seoul National University

Fall 2026

Overview

This course is designed for graduate students who want to understand, and contribute to, the two frontiers that are rapidly converging in modern AI: *multimodal foundation models* that perceive, reason, and generate across vision, language, and audio, and *physical world models* that represent, simulate, and predict the 3D/4D physical world under actions. About half of the semester is devoted to intensive lectures, given as two series: one on multimodal foundation models, covering multimodal LLMs and multimodal generative AI, and one on 3D/4D representation and generation together with a review of the recent world-model literature. The rest of the semester is a paper-driven seminar in which students present and critique papers selected from a topic pool proposed by the instructors, followed by an open discussion in which the whole class brainstorms possible future research directions. The two components are interleaved over the semester rather than strictly separated into a first and a second half. Prior knowledge of machine learning and deep learning is assumed.

1 Target Audience

- Graduate students with prior knowledge in machine learning and deep learning who want to specialize in multimodal foundation models and physical world models.

2 Course Objectives

- Understand the design principles of multimodal foundation model theories and applications: architecture, training recipe, data, and evaluation.
- Understand the design principles of multimodal generative model theories and applications expanding to 3D/4D and video generation.
- Critically review and discuss the emerging literature on multimodal foundation models and physical world models.

3 General Information

3.1 Class information

- Friday 1:00 PM – 3:50 PM, Sep. 4th to Dec. 18th
- Bldg. 302, Room 106
- Credit 3-3-0

3.2 Office hour

- Friday 10:00 AM, AIIS Bldg. 942, Room 208
- Reservation only

- Sangdoo Yun (sangdoo.yun@navercorp.com), Jin-Hwa Kim (jnhwkim@snu.ac.kr)

4 Prerequisites

- Engineering Mathematics, Discrete Mathematics, Introduction to Artificial Intelligence, Computer Graphics, Machine Learning

5 Grade

- Attendance (10%)
  - Students who are absent for over 1/3 of the class will receive a grade of F for the course. (Exceptions can be made when the course instructor deems the cause of absence unavoidable.)
- Attitude (30%)
  - Asking a thoughtful question (as determined by the instructor or invited speaker) earns 5 (good) or 10 (excellent) points.
- Presentation (30%)
- Final Report (30%)

6 Schedule

The following is a tentative schedule for the course:

Week	Date	Topic	Note
1	Sep 4	Introduction	JH & SD
2	Sep 11	Multimodal Foundation Model I	SD
3	Sep 18	Multimodal Foundation Model II	SD
4	Sep 25	Multimodal Foundation Model III (Online)	SD
5	Oct 2	Multimodal Generation I	JH
6	Oct 9	Multimodal Generation II (Online)	JH
7	Oct 16	Multimodal Generation III	JH
8	Oct 23	Paper Review and Discussion Session 1	–
9	Oct 30	Paper Review and Discussion Session 2	–
10	Nov 6	Paper Review and Discussion Session 3	–
11	Nov 13	Paper Review and Discussion Session 4	–
12	Nov 20	Paper Review and Discussion Session 5	–
13	Nov 27	Paper Review and Discussion Session 6	–
14	Dec 4	Paper Review and Discussion Session 7	–
15	Dec 11	NeurIPS 2026 (No Class)	–
16	Dec 18	Paper Review and Discussion Session 8	–

Note: The course content and sequence may be adjusted based on the duration of the lecture and the interests of the students. Additional readings, assignments, and projects may also be included to enhance the learning experience.