

4190.773: Multimodal Generative AI Theories and Applications

Jin-Hwa Kim^{1,2} Sangdoo Yun^{1,2}
¹NAVER AI Lab ²Seoul National University

Fall 2025

Overview

This course is designed for graduate students interested in developing theories and applications of multimodal generative AI. The course will cover core concepts and fundamental theories that are applied in various enterprise applications. Assuming prior knowledge in machine learning and related topics, the course will explore multimodal generative AI from vision and language to audio and speech using deep neural networks and the techniques used in deep learning literature. Furthermore, an emerging topic of neural graphics will also be discussed, focusing on studies on the NeRFs and 3DGS in the aspect of multi-view representation learning.

1 Target Audience

- Graduate students with prior knowledge in machine learning and related courses who are looking to specialize or broaden in multimodal deep learning.

2 Course Objectives

- Gain a deep understanding of multimodal deep learning theories and applications
- Explore various modalities such as vision, language, audio, and speech in deep learning
- Stay updated with emerging topics in neural graphics, specifically NeRFs and 3DGS

3 General Information

3.1 Class information

- Friday 1:00 PM - 3:50 PM, Sep. 5th to Dec. 12th
- Bldg. 302, Room 107
- Credit 3-3-0

3.2 Office hour

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

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

4 Grade

- Attendance (10%)
 - Students who miss more than one-third of the classes without a valid excuse will receive 0 points. Otherwise, full credit (10 points) will be given.
- Attitude (10%)
 - Asking a thoughtful question (as determined by the instructor or invited speaker) earns 1 (good) or 2 (excellent) points.
 - Giving a short presentation (12 minutes + 3 minutes Q&A) on a topic related to multimodal generative AI earns 1–5 points. Up to 12 presentation opportunities are available in advance of the scheduled lectures. First come, first served. Students can also earn points by asking questions during these presentations.
 - The base score is 5 points and may be deducted at the lecturer’s discretion.
- Midterm exam (40%)
- Final exams (40%)

5 Schedule

The following is a tentative schedule for the course:

Week	Date	Topics	Lecturer
1	5 Sep	Introduction	JH & SD
2	12 Sep	Multimodal Representation Learning	JH
3	19 Sep	Multimodal Foundation Models I	SD
4	26 Sep	Multimodal Foundation Models II	SD
5	3 Oct	National Foundation Day (No Class)	-
6	10 Oct	Multimodal Generation I	Junho Kim (Invited)
7	17 Oct	Multimodal Foundation Models III	SD
8	24 Oct	Midterm Exam	-
9	31 Oct	Neural Graphics I	JH
10	7 Nov	Multimodal Generation II	Gayoung Lee (Invited)
11	14 Nov	Neural Graphics II	JH
12	21 Nov	Neural Graphics III	JH
13	28 Nov	Trustworthy AI	SD
14	5 Dec	Speech Generation	Eunwoo Song (Invited)
15	12 Dec	Human-Computer Interaction	Youngho Kim (Invited)
16	19 Dec	Final Exam	-

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.