

MULTIMODAL MACHINE LEARNING – FALL 2026

Time: Tuesdays and Thursday, 9:30am-10:50am

Classroom: POS A35

Zoom: Only by request, 2 days in advance.

Canvas: <https://canvas.cmu.edu/courses/55519>

Piazza: <https://piazza.com/cmu/fall2026/11777a/info>

Instructor: Prof. Louis-Philippe Morency, morency@cs.cmu.edu

Teaching assistants:

- Kashish Gandhi, kgandhi2@andrew.cmu.edu
- Manu Gaur, mgaur@andrew.cmu.edu
- Shruti Jain, spjain@andrew.cmu.edu
- Yousha Mahamuni, ymahamun@andrew.cmu.edu
- Nikita Malik, nikitama@andrew.cmu.edu
- Aaditya Vikram, aadityas@andrew.cmu.edu
- Koushik Viswanadha, koushikv@andrew.cmu.edu
- Danqing Wang, danglingw@andrew.cmu.edu
- Yuteng Zhang, yutengz@andrew.cmu.edu

Course Description

Multimodal machine learning (MMML) is a vibrant interdisciplinary research area that advances core goals of artificial intelligence by integrating and modeling multiple communicative modalities, including language, speech, vision, and other sensory signals. Originating from early work in audio-visual speech recognition and evolving through modern language-and-vision systems such as image and video captioning, MMML addresses unique challenges arising from heterogeneous data sources and complex interdependencies between modalities.

This course introduces the fundamental mathematical and algorithmic concepts in machine learning and deep learning that underpin the **six central challenges of multimodal machine learning**:

1. Representation
2. Alignment
3. Reasoning
4. Generation
5. Transference
6. Quantification

Students will study both foundational and modern approaches to these challenges, including multimodal autoencoders, deep canonical correlation analysis, multimodal tensor fusion, multi-kernel learning, attention mechanisms, multimodal recurrent models, multimodal transformers, neuro-symbolic methods, mutual information-based objectives, and multimodal graph neural networks.

The course also examines state-of-the-art probabilistic models and computational algorithms through readings of recent research literature, with attention to open problems and emerging

directions in the field. Applications discussed throughout the course include multimodal affect recognition, language grounding, language-vision navigation, image and video captioning, and cross-modal multimedia retrieval.

Preparation

This course is a graduate-level Carnegie Mellon University for-credit course (and is not an executive education or professional development type of offering). You are required to participate fully, attend class sessions, collaborate within teams, and to complete assignments on time.

Students should have pre-requisite background knowledge in probability, statistics, and linear algebra. Prior knowledge of programming in Python and deep learning is recommended.

Course format

In this course, students are expected to attend in-person the lectures, scheduled on Tuesdays and Thursdays. Each lecture will focus on specific theoretical and mathematical concepts related to multimodal machine learning. Students will have individual as well as group assignments in the form of readings and project work. Reading assignments are designed to complement the lecture materials and showcase recent state-of-the-art research. Reading material will be based on published technical papers available via the ACM/IEEE/ Springer digital libraries or freely available online (e.g., arxiv.org). CMU students have free access to these digital archives. Team project work includes regular reports that are due at intervals throughout the semester (see more details below).

Grading Scale

- Lecture participation (16%)
- Reading assignments (12%)
- Course project assignments
 - Project preferences and pre-proposal (2%)
 - First project assignments (10%)
 - Proposal presentation (5%)
 - Second project assignment (10%)
 - Mid-term report (10%) and presentation (5%)
 - Final report (20%) and presentation (10%)

Course Schedule

Exact weekly topics may change based on a number of factors (e.g., emerging news in the field, time restrictions, interest). ** Weeks are marked with (read) when a reading assignment is planned and marked with (proj) when a project assignment is due.

Classes	Tuesday Lectures	Thursday Lectures
Week 1 8/25 & 8/27	Course introduction • Multimodal core challenges • Course syllabus	Multimodal applications and datasets • Designing research questions • Datasets and research task
Week 2 9/1 & 9/3 <i>Read due: 9/4</i>	Unimodal representations • Language representations • Self-attention transformer models	Unimodal representations • Visual and signal representations • Dimensions of heterogeneity
Week 3 9/8 & 9/10 <i>Proj. Due: 9/9</i> <i>Read due: 9/12</i>	Multimodal representations • Cross-modal interactions • Multimodal fusion	Multimodal representations • Coordinated representations • Multimodal fission
Week 4 9/15 & 9/17 <i>Proj. due: 9/20</i>	Multimodal alignment and grounding • Discrete and continuous alignment • Multimodal grounding	Transference • Cross-modal distillation, co-learning • Pretraining, fine-tuning, post-training
Week 5 9/22 & 9/24 <i>Read due: 9/25</i>	<i>Project presentations (proposal)</i>	<i>Project presentations (proposal)</i>
Week 6 9/29 & 10/1 <i>Proj. due: 10/4</i>	Visual-Language Models • Visual-language alignment • Multimodal LLMs	Video and Unified Spaces • Binding multiple modalities • Video transformers
Week 7 10/6 & 10/8 <i>Read due: 10/9</i>	Guest lecture	Guest lecture – Alex Wilf • Web agents – LLMs and MLLMs • Self-learning and verification
Week 8 10/13 & 10/15	<i>Fall Break – No lectures</i>	
Week 9 10/20 & 10/22 <i>Proj. due: 10/25</i>	Multimodal Reasoning • Modular and neuro-symbolic models • Graph and structured models	Multimodal Reasoning • Test-time computation • Multimodal chain-of-thought
Week 10 10/27 & 10/29	<i>Project presentations (midterm)</i>	<i>Project presentations (midterm)</i>
Week 11 11/3 & 11/5 <i>Read due: 11/6</i>	<i>Democracy Day – No Class –</i>	Multimodal Generation • Autoregressive generation •
Week 12 11/10 & 11/12 <i>Read due: 11/13</i>	Multimodal Generation • Diffusion and flow matching • Omni any-to-any models	Visual-Language-Action • Embodied representations • VLA foundation models
Week 13 11/17 & 11/19	Quantification and Bias • Heterogeneity and bias • Evaluation and learning	New Research Directions Recent research in multimodal ML
Week 14 11/24 & 11/26	<i>Thanksgiving Week – No Class –</i>	<i>Thanksgiving Week – No Class –</i>
Week 15 12/1 & 12/3 <i>Proj. due: 12/6</i>	<i>Project presentations (final)</i>	<i>Project presentations (final)</i>

Projects

The goal of the course project is to experiment with state-of-the-art multimodal algorithms and computational models. Students have flexibility in the selection of their project topic. The project should be directly aligned with the course content and include at the minimum two modalities, preferably language and vision. At the beginning of the semester, the instructor will propose a set of research problems and datasets which can be used for the course projects.

Students should create teams between 4 to 5 students preferably (special approval will be required for larger teams; no smaller teams will be allowed). The size and depth of the project should be adjusted to reflect the size of the team. Each team is required to create a code repository (github) for their project. All project members should be included in this project and should actively use it. It is important that all team members participate equally in the project. Each project assignment report will need to outline the tasks of each student. If any team member has concern in the participation level of other members, they should contact the instructor and/or TA as promptly as possible.

Timeline and assignments:

- **Dataset and project preferences** (Due on Tuesday 9/1 at 8pm ET) – Let us know your preferences for the datasets that you would like to use for the course project. This will help with the team matching process.
- **Project Pre-proposal** (Due on Wednesday 9/9 at 8pm ET) – You should have selected your teammates, dataset, and task. Submit a 1-page pre-proposal plan.
- **First assignment** (Due Sunday 9/20 at 8pm ET and presentations are scheduled Tuesday 9/22 and Thursday 9/24) This assignment requires a good literature review on the topic of your proposed project and some initial design for the course project, including the selection of the dataset and initial draft for the research ideas.
- **Second assignment** (Due Sunday 10/4 at 8pm ET) – This assignment focuses on a better understanding of the data itself, including some unimodal and multimodal data analyses.
- **Midterm assignment** (Report is due Sunday 10/25 at 8pm ET and presentations are scheduled Tuesday 10/27 and Thursday 10/29) – Students are asked to implement and evaluate state-of-the-art baseline models on their project dataset and perform error analysis.
- **Final assignment** (Presentations are planned Tuesday 12/1 and Thursday 12/3; reports are due Sunday 12/6 at 8pm ET) – Students should explore new ideas to model their multimodal research project.

(See Canvas and Piazza for additional information)

Policies

Attendance Policy. Since you signed up for this class, you are expected to attend all lectures. However, we all know that sometimes things get in the way.

Each class we will require you to check in, but you will get 4 unexcused absences without a warning (i.e., absences that you don't have to explain why you're not there). You do get as many excused absences (e.g., due to sickness) as needed per CMU university policy. Note that for longer-term absences and accommodations (e.g., long term sickness, taking care of family members, etc.) you will need to contact CMU's [Office of Disability Resources](#).

Also note, if you arrive 5+ minutes late you will be marked as absent for that class.

Late days policy. There are no late days for any of the assignments, whether it be reading or project assignment deadlines. For reading assignments, any lateness will result in a zero for that assignment. For project assignments, for every day that you submit your homework late, you will receive a deduction proportional to the lateness (10% per late day; if you're 2 days late, you will lose 20% of your grade). You have a 30-minute grace period before the lateness deduction will be applied (i.e., you can submit up to 30 minutes after the deadline without any penalty, but any minute after will result in 10% deduction).

However, if you need an extension, you are free to ask the instructor and we will review on a case-by-case basis, but you must email at least 24h before the deadline.

Academic honesty. Reading assignments are to be completed individually - everything you turn in must be your own work, and you must cite resources that you used to learn about the problem. The project is to be completed by a team. You are encouraged to use existing NLP components in your project; you must acknowledge these appropriately in the documentation. Suspected violations of academic integrity rules will be handled in accordance with the [CMU guidelines on collaboration and cheating](#).

AI-assisted writing policy. You are allowed to use LLMs and generative AI in your assignments and projects, unless specified otherwise. However, you are responsible for all the AI's outputs as if they were your own, so make sure that you verify the outputs, make sure they are detailed enough, etc. Also, **You must disclose in your writeups how you used GenAI in your workflows** in a detailed manner (e.g., help with grammar correction, automating implementation of certain functions, turning bullets into prose, finding relevant papers for topic X, etc.); do not simply say things like help with coding, help with writing as that is not specific enough.

Accommodations for students with disabilities. If you have a disability and have an accommodations letter from the Disability Resources office, we encourage you to discuss your accommodations and needs with the instructors as early in the semester as possible. We will work with you to ensure that accommodations are provided as appropriate. If you suspect that you may have a disability and would benefit from accommodations but are not yet registered with the [Office of Disability Resources](#), we encourage you to contact them via their website.

Note to Students

Statement on Student Wellness. As a student, you may experience a range of challenges that can interfere with learning, such as strained relationships, increased anxiety, substance use, feeling down, difficulty concentrating and/or lack of motivation. This is normal, and all of us benefit from support during times of struggle. There are many helpful resources available on campus and an important part of a healthy life is learning how to ask for help. Asking for support sooner rather than later is almost always helpful. CMU services are available free to students, and treatment does work. You can learn more about confidential mental health services available on campus through [Counseling and Psychological Services \(CaPS\)](#). Support is always available (24/7) at: 412-268-2922.

Every individual must be treated with respect. In this class, every individual will and must be treated with respect. The ways we are diverse are many and are fundamental to building and maintaining an equitable and an inclusive campus community. These include but are not limited to: race, color, national origin, caste, sex, disability (visible or invisible), age, sexual orientation, gender identity, religion, creed, ancestry, belief, veteran status, or genetic information.

Research shows that greater diversity across individuals leads to greater creativity in the group. We at CMU work to promote diversity, equity and inclusion not only because it is necessary for excellence and innovation, but because it is just. Therefore, while we are imperfect, we ask you all to fully commit to work, both inside and outside of our classrooms to increase our commitment to build and sustain a campus community that embraces these core values. It is the responsibility of each of us to create a safer and more inclusive environment. Incidents of bias or discrimination, whether intentional or unintentional in their occurrence, contribute to creating an unwelcoming environment for individuals and groups at the university. If you experience or observe unfair or hostile treatment on the basis of identity, we encourage you to speak out for justice and offer support in the moment and/or share your experience using the following resources:

- [Center for Student Diversity and Inclusion: csdi@andrew.cmu.edu](mailto:csdi@andrew.cmu.edu), (412) 268-2150
- [CMU anonymous reporting hotline](#), (844) 587-0793