

# MULTIMODAL MACHINE LEARNING – FALL 2023

[Description](#) | [Schedule](#) | [Grades](#) | [Bibliography](#)

**Instructor:** Prof. Louis-Philippe Morency, [morency@cs.cmu.edu](mailto:morency@cs.cmu.edu)

**Co-lecturer:** Paul Liang, [pliang@cs.cmu.edu](mailto:pliang@cs.cmu.edu)

## Teaching assistants:

- Syeda Akter [sakter@andrew.cmu.edu](mailto:sakter@andrew.cmu.edu)
- Mehul Agarwal [mehula@andrew.cmu.edu](mailto:mehula@andrew.cmu.edu)
- Aditya Rathod [arathod@andrew.cmu.edu](mailto:arathod@andrew.cmu.edu)
- Soham Dinesh Tiwari [sohamdit@andrew.cmu.edu](mailto:sohamdit@andrew.cmu.edu)
- Haofei Yu [haofey@andrew.cmu.edu](mailto:haofey@andrew.cmu.edu)

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

**Classroom:** MM A14

**Recommended preparation:** This is a graduate course designed primarily for PhD and research master students at LTI, MLD, CSD, HCII, and RI; others, for example (undergraduate) students of CS or from professional master programs, are advised to seek prior permission of the instructor. It is required for students to have taken an introduction machine learning course such as 10-401, 10-601, 10-701, 11-663, 11-441, 11-641 or 11-741. Prior knowledge of deep learning is recommended. Students should have proper academic background in probability, statistics, and linear algebra. Programming knowledge in Python is also strongly recommended.

## Introduction and Purposes

Multimodal machine learning (MMML) is a vibrant multi-disciplinary research field which addresses some of the original goals of artificial intelligence by integrating and modeling multiple communicative modalities, including linguistic, acoustic, and visual messages. With the initial research on audio-visual speech recognition and more recently with language & vision projects such as image and video captioning, this research field brings some unique challenges for multimodal researchers given the heterogeneity of the data and the contingency often found between modalities. This course will teach fundamental mathematical concepts related to MMML including multimodal alignment and fusion, heterogeneous representation learning and multi-stream temporal modeling. We will also review recent papers describing state-of-the-art probabilistic models and computational algorithms for MMML and discuss the current and upcoming challenges.

The course will present the fundamental mathematical concepts in machine learning and deep learning relevant to the six main challenges in multimodal machine learning: (1) representation, (2) alignment, (3) reasoning, (4) generation, (5) transference and (6) quantification. These include, but not limited to, multimodal transformers, neuro-symbolic models, multimodal tensor fusion, mutual information and multimodal graph networks. The course will also discuss many of the recent applications of MMML including multimodal affect recognition, multimodal language grounding and language-vision navigation.

## Course format

Lectures will be happening in-person on Tuesdays and Thursdays at 9:30am. Each lecture will focus on a specific mathematical concept related to multimodal machine learning. These lectures will be given by the course instructor, a guest lecturer or a TA.

**Canvas** We will be using Canvas for all course assignments. Students should be automatically added to the course Canvas site.

**Google Forms** For some course assignments (e.g., the lecture highlights), Google Forms will also be used. The link to each individual form will be communicated to students using Piazza.

## Course Material

### Required:

- 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). All CMU students have already free access to these digital archives.
- For project assignments, previous experience in Python programming is expected

## Project Assignments and Timeline

(See Piazza for additional information)

- **Dataset and project preferences** (Due on Tuesday 9/5 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 9/13 at 8pm ET) – You should have selected your teammates, dataset, and task. Submit a 1-page pre-proposal plan.
- **First assignment** (Due Sunday 9/24 at 8pm ET) This assignment requires a good literature review on the topic of your proposed project.
- **Second assignment** (Due Sunday 10/8 at 8pm ET) – This assignment focuses on unimodal representations and dataset analyses.
- **Midterm assignment** (Report is due Sunday 10/29 at 8pm ET and presentations are scheduled Tuesday and Thursday 10/31 and 11/2) – 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 12/5 and Thursday 12/7; reports are due Monday 12/10 at 8pm ET) – Students should explore new ideas to model their multimodal research project.

## Course Topics

**\*\* Exact topics may change based on student interests and time restrictions. \*\* 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
<b>Week 1</b> 8/29 & 8/31	<b>Course introduction</b> <ul style="list-style-type: none"> <li>• Multimodal core challenges</li> <li>• Course syllabus</li> </ul>	<b>Multimodal applications and datasets</b> <ul style="list-style-type: none"> <li>• Research tasks and datasets</li> <li>• Team projects</li> </ul>
<b>Week 2</b> 9/5 & 9/7 <i>Read due: 9/9</i>	<b>Unimodal representations</b> <ul style="list-style-type: none"> <li>• Dimensions of heterogeneity</li> <li>• Visual representations</li> </ul>	<b>Unimodal representations</b> <ul style="list-style-type: none"> <li>• Language representations</li> <li>• Signals, graphs and other modalities</li> </ul>
<b>Week 3</b> 9/12 & 9/14 <i>Read due: 9/16</i> <i>Proj. Due: 9/13</i>	<b>Multimodal representations</b> <ul style="list-style-type: none"> <li>• Cross-modal interactions</li> <li>• Multimodal fusion</li> </ul>	<b>Multimodal representations</b> <ul style="list-style-type: none"> <li>• Coordinated representations</li> <li>• Multimodal fission</li> </ul>
<b>Week 4</b> 9/19 & 9/21 <i>Proj. due: 9/24</i>	<b>Multimodal alignment and grounding</b> <ul style="list-style-type: none"> <li>• Explicit alignment</li> <li>• Multimodal grounding</li> </ul>	<b>Alignment and representations</b> <ul style="list-style-type: none"> <li>• Self-attention transformer models</li> <li>• Masking and self-supervised learning</li> </ul>
<b>Week 5</b> 9/26 & 9/28 <i>Read due: 9/30</i>	<b>Multimodal transformers</b> <ul style="list-style-type: none"> <li>• Multimodal transformers</li> <li>• Video and graph representations</li> </ul>	<b>Multimodal reasoning</b> <ul style="list-style-type: none"> <li>• Structured and hierarchical models</li> <li>• Memory models</li> </ul>
<b>Week 6</b> 10/3 & 10/5 <i>Proj. due: 10/8</i>	<b><i>Project hours</i></b>	<b><i>Multimodal language grounding</i></b> <ul style="list-style-type: none"> <li>• Guest lecturer: Jack Hessel</li> <li>• Vision, language and grounding</li> </ul>
<b>Week 7</b> 10/10 & 10/12 <i>Read due: 10/14</i>	<b>Multimodal interaction</b> <ul style="list-style-type: none"> <li>• Reinforcement learning</li> <li>• Discrete structure learning</li> </ul>	<b>Multimodal inference</b> <ul style="list-style-type: none"> <li>• Logical and causal inference</li> <li>• External knowledge</li> </ul>
<b>Week 8</b> 10/17 & 10/19	<b><i>Fall Break – No lectures</i></b>	
<b>Week 9</b> 10/24 & 10/26 <i>Proj. due: 10/29</i>	<b>Multimodal generation</b> <ul style="list-style-type: none"> <li>• Translation, summarization, creation</li> <li>• Generative models: VAEs</li> </ul>	<b>New generative models</b> <ul style="list-style-type: none"> <li>• GANs and diffusion models</li> <li>• Model evaluation and ethics</li> </ul>
<b>Week 10</b> 10/31 & 11/2	<b><i>Project presentations (midterm)</i></b>	<b><i>Project presentations (midterm)</i></b>
<b>Week 11</b> 11/7 & 11/9 <i>Read due: 11/12</i>	<b><i>Democracy Day – No Class –</i></b>	<b>Transference</b> <ul style="list-style-type: none"> <li>• Modality transfer and co-learning</li> <li>• Self-training and multitask learning</li> </ul>
<b>Week 12</b> 11/14 & 11/16 <i>Read due: 11/21</i>	<b>Quantification</b> <ul style="list-style-type: none"> <li>• Heterogeneity and interactions</li> <li>• Biases and fairness</li> </ul>	<b>New research directions</b> <ul style="list-style-type: none"> <li>• Recent research in multimodal ML</li> </ul>
<b>Week 13</b> 11/21 & 11/23	<b><i>Thanksgiving Week – No Class –</i></b>	
<b>Week 14</b> 11/28 & 11/30	<b>Guest lecture</b>	<b>Guest lecture</b>
<b>Week 15</b> 12/5 & 12/7 <i>Proj. due: 12/10</i>	<b><i>Project presentations (final)</i></b>	<b><i>Project presentations (final)</i></b>

## Grades

- **Grading breakdown**
  - Lecture participation and highlights 16%
  - Reading assignments 12%
  - Course project assignments
    - Project preferences and pre-proposal 2%
    - First project assignments 10%
    - Second project assignment 10%
    - Mid-term report and presentation 20%
    - Final report and presentation 30%
- **Lecture participation and highlights**
  - Lectures are expected to be attended live, in-person. Live attendance allows students to participate in the discussions and ask any clarification questions.
  - During the lecture (or soon after), students are expected to fill a form where they include, they summarize their main takeaways from the lecture (aka, highlights).
  - The form should be submitted the same day as the lecture, before 11:59pm ET. For example, if the lecture was scheduled Tuesday, then the highlight form is due on the same Tuesday at 11:59pm ET.
  - Students need to use the provided online template for the highlight form. This form was designed for two main purposes: (1) help students for taking active notes during lectures, and (2) offer students the opportunity to ask questions about the content of the lectures.
    - The lecture is split in three 30-minute segments (the last segment may be shorter).
    - For each segment, students are asked to include a short statement summarizing the main points of the past segment.
    - Students are also asked to include their main take-away messages for the whole lecture.
    - Optionally students can also write down a question (with corresponding slide number) related to the segment.
  - The student's questions will be reviewed by TAs and instructor. The most popular questions will be answered through discussion board, or with extra information during the following lecture. Students are always welcome to post questions directly on the discussion board at any time if they would like clarifications or have a follow-up question.
  - These highlight forms will not be required for the first week, for the Thanksgiving week and project presentation weeks. Also, no forms are expected for "project hours" lectures which are organized to help students with project assignments. We expect about 16 lectures where highlight forms need to be submitted.
  - Each submitted form will be graded for 1.0 point.
- **Reading assignments**
  - Reading assignments are designed to complement the lectures and showcase recent state-of-the-art research. Most reading assignments will consist of multiple research

papers, sometimes accompanied by optional readings. The list of research papers will be released at the latest on the Monday of each week.

- To encourage exchange of ideas and knowledge between students, each student will be part of one study group. A study group consists of 8-10 students. These groups will be randomly created, to encourage diversity in these groups. Each study group will have its own discussion forum to ask questions and share ideas.
- The reading assignments consist of two main parts: (a) selecting one reading paper and creating a quality summary of the paper, and (b) reading the summary from other papers that were not read by the student, and submit a discussion essay describing the main differences and follow-up ideas.
  - For each reading assignment, each student is required to read only one research paper (out of all assigned papers). Students need to write a summary statement for their paper and post it in the discussion forum before Friday 8pm ET. These summary posts will allow other study group members to learn about the papers they did not read directly, and possibly ask follow-up questions.
  - Following this Friday deadline, students are expected to read at least one summary for each paper they did not originally read and write a follow-up post to discuss similarities and differences with the paper they read and suggest some follow-up ideas. One follow-up post should be created for each paper the student did not read. For most reading assignments, the total number of paper options is 3, making the number of follow-up posts equal to 2.
- A total of 6 reading assignments are expected during the semester.
  - A larger number of assignments is concentrated at the beginning of the semester to help students learn about the basic concepts related to multimodal machine learning.
  - Reading assignments are spaced out in later parts of the semester to help students focus on the project assignments.
  - These reading assignments are planned for the following weeks:
    - Week 2, Week 3, Week 5, Week 7, Week 11 and Week 12.
- Each reading assignment is worth 2.0 points: 1.0 point for the paper summary and 1.0 point for the short discussion essay.
- **Course project assignments**
  - The goal of the course project is to experiment with state-of-the-art multimodal algorithms and computational models.
  - Students should create teams between 3 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 to the project. The first project assignment and follow-up reports (midterm and final reports) 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.
  - 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.

- **Pre-proposal:** We ask students to prepare a pre-proposal early in the semester to them establish their research topic for the course project. The pre-proposal will also help with team formation, in the rare eventuality when students are still looking for teammates.
- **First project assignment:** The first project assignment has three main goals: (1) describing the general problem and dataset that will be explored as part of the course project, (2) summarize the related work on this topic, and (3) propose some ideas on how this research could be explored forward.
- **Second project assignment:** The second project assignment has the main goal to start exploring the project dataset. To start this exploration, we ask teams to perform some unimodal analyses on their multimodal dataset and research problem.
- **Midterm project assignment:** The midterm project assignment is designed to implement multimodal baseline models and perform some error analysis on these results. This assignment has two components: written report and oral presentation. The written report is submitted first and should also include some updated research ideas on how to explore this research topic. The oral presentations will be done soon after submission of the midterm reports to gather feedback from other classmates about their course projects and research ideas.
- **Final project assignment:** Using the same dataset selected for the midterm report, the final project assignment is designed to explore new research ideas. This assignment is not graded based on better dataset performance, but instead on the exploration of new ideas. In other words, better accuracy results will not mean better course grade. Students are encouraged to explore new research directions. The final project assignment consists of a written report and an in-person poster presentation. The poster presentations may be opened to the general CMU population, if the COVID situation allows it.

### **Notes about late submissions**

In general, submitting assignments on time lets the instructional team provide feedback in a more timely and efficient manner. Timely submissions are particularly important for assignments with discussions and peer feedback, such as the reading assignments and the project assignments. Also, it is expected that students will attend the lectures in person. Live attendance allows students to participate in the discussions and ask any clarification questions.

**Medical-related absences** If for medical reason you require some extra time for an assignment or may not be able to attend the lecture in person, please contact instructor and the TAs as soon as possible (the best option is usually via Piazza) and we will help define a new plan that aligns with your constraints.

**Late submission wildcards** We offer students and project teams some late submission wildcard to help deal with potential overlaps with other courses or research deadlines. The details are expressed below.

- Reading assignments and lecture highlights
  - Each student will receive six (6) late submission wildcards, to be used individually.

- Each wild card can be used to extend the deadline up to 24 hours from the original time.
- No partial credit for the wild card (e.g., it is not possible to use only half of the card, with two times 12 hours).
- These wildcards can be used together (for a total of 48 hours), or separately (2 separate extensions of 24 hours).
- There is no need to send a note via Piazza about these wildcards. We will automatically use your wildcards your first two late submissions.
- For any other late submission (beyond the two wildcards) then 0.5 points will be deducted.
- If you must submit beyond 72 hours past the due date, please contact instructor and the TAs as soon as possible so we can properly plan.
- Project assignments
  - Each team will get two (2) wild cards, to be used with any of the project assignment deadlines.
  - Each wild card allows the team to submit their assignment late for up to 24 extra hours.
  - These wild cards can be used together (for a total of 48 hours), or separately (2 separate extensions of 24 hours).
  - Each wild card can be used for any of these 4 deadlines:
    - First assignment deadline
    - Second assignment deadline
    - Midterm report deadline
    - Final report deadline
  - No partial credits for the wild cards (e.g., you cannot use only 40% of a wild card).
  - Each team needs to send a message to instructors BEFORE the deadline to notify TAs and instructor about the intent to use a wild card (or two).

### **Accommodations for Students with Disabilities**

If you have a disability and have an accommodations letter from the Disability Resources office, I encourage you to discuss your accommodations and needs with me as early in the semester as possible. I 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, I encourage you to contact them at [access@andrew.cmu.edu](mailto:access@andrew.cmu.edu).

### **Statement on Student Wellness**

This semester is unlike any other. We are all under a lot of stress and uncertainty at this time. Attending Zoom classes all day can take its toll on our mental health. Make sure to move regularly, eat well, and reach out to your support system or me ([morency@cs.cmu.edu](mailto:morency@cs.cmu.edu)) if you need to. We can all benefit from support in times of stress, and this semester is no exception.

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. These mental health concerns or stressful events may diminish your academic performance and/or reduce your ability to participate in daily activities. CMU services are available, and treatment does work. You can learn more about confidential mental health services available on campus at: <http://www.cmu.edu/counseling/>. Support is always available (24/7) from Counseling and Psychological Services: 412-268-2922.

## Diversity statement

**Every individual 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, sex, disability, age, sexual orientation, gender identity, religion, creed, ancestry, belief, veteran status, or genetic information. We at CMU, will 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 all need 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 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, [www.cmu.edu/student-diversity](http://www.cmu.edu/student-diversity)
- Report-It online anonymous reporting platform: [www.reportit.net](http://www.reportit.net) username: tartans password: plaid

All reports will be acknowledged, documented, and a determination will be made regarding a course of action. All experiences shared will be used to transform the campus climate to be more equitable and just.

## Bibliography

**Reading lists from previous courses** are available on Piazza:

<https://piazza.com/cmu/fall2018/11777/resources>

<https://piazza.com/cmu/fall2019/11777/resources>

<https://piazza.com/cmu/fall2020/11777a/resources>

<https://piazza.com/cmu/fall2021/11777/resources>

<https://piazza.com/cmu/fall2022/11777/resources>

The reading list for Fall 2023 semester will also be shared directly with students and posted on the course website.