Language Technologies Institute

Carnegie Mellon University

Advanced Topics in Multimodal Machine Learning (11-877)

Lecture 1: Introduction

Louis-Philippe Morency and Paul Liang

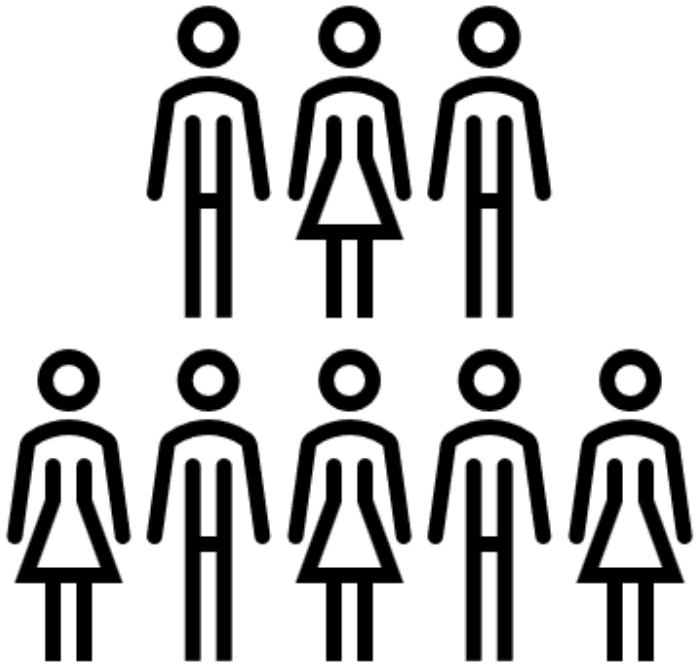
Your Teaching Team This Semester (11-877, Spring 2023)

Louis-Philippe Morency
morency@cs.cmu.edu
Course instructor

Paul Liang
pliang@cs.cmu.edu
Course instructor

Catherine (Yun) Cheng
yuncheng@andrew.cmu.edu
Teaching Assistant

Time for Introductions!



Your name, department and programs

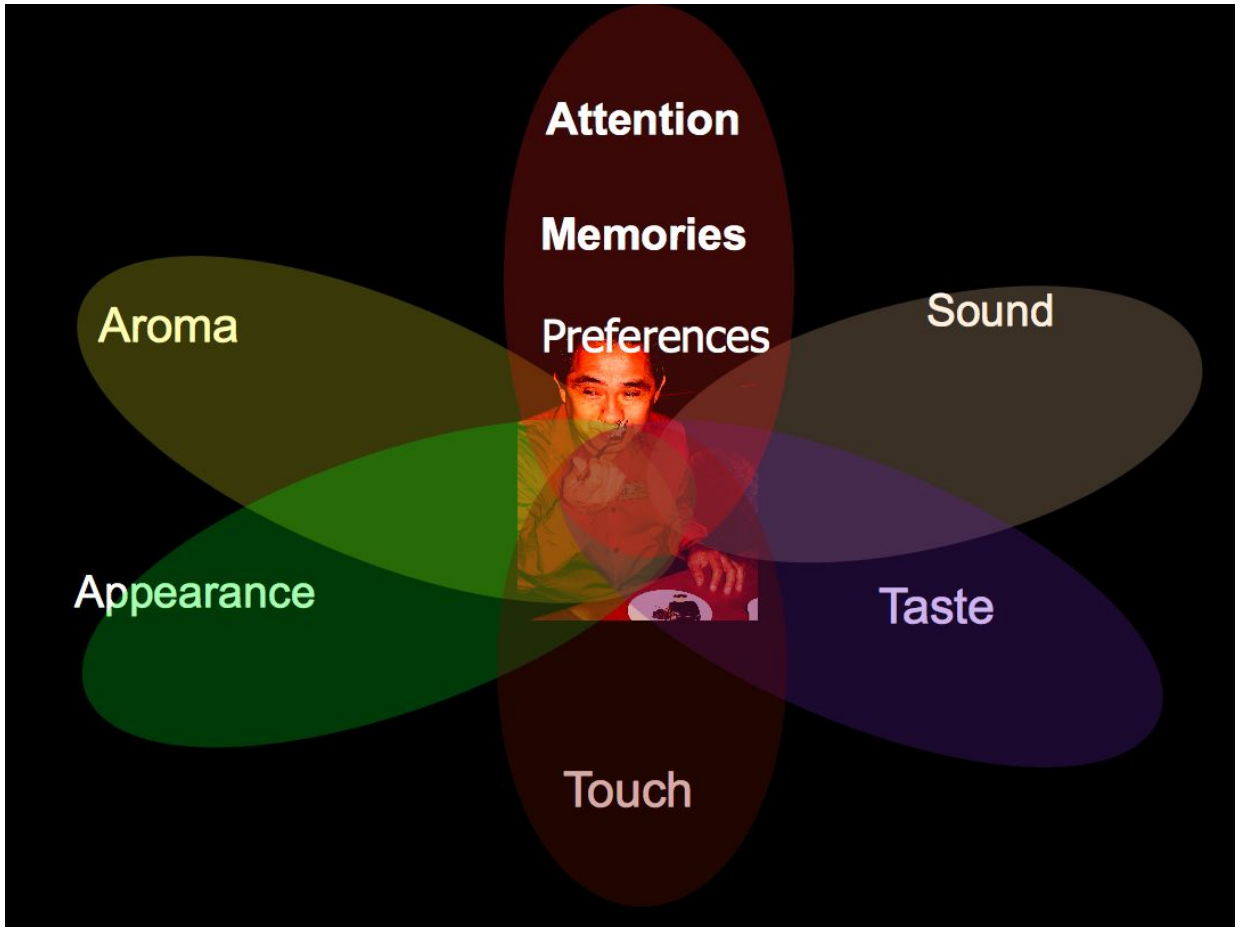
Your favorite modality(ies)!

Previous research experience in multimodal

Why are you interested in this course?

What is Multimodal?

What is Multimodal?

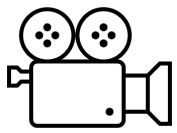


Sensory Modalities

What is a Modality?

Modality

Modality refers to the way in which something expressed or perceived.



from a sensor

Raw Modalities
(closest from sensor)

Abstract Modalities
(farthest from sensor)

Examples:

Speech signal	Language	Sentiment intensity
Image	Detected objects	Object categories

What is Multimodal?

A dictionary definition…

Multimodal: with multiple modalities

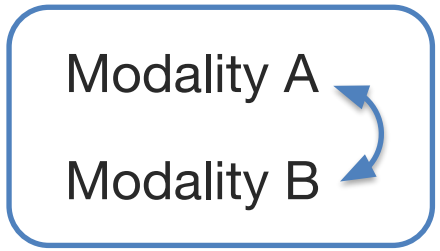
A research-oriented definition…

***Multimodal* is the science of**

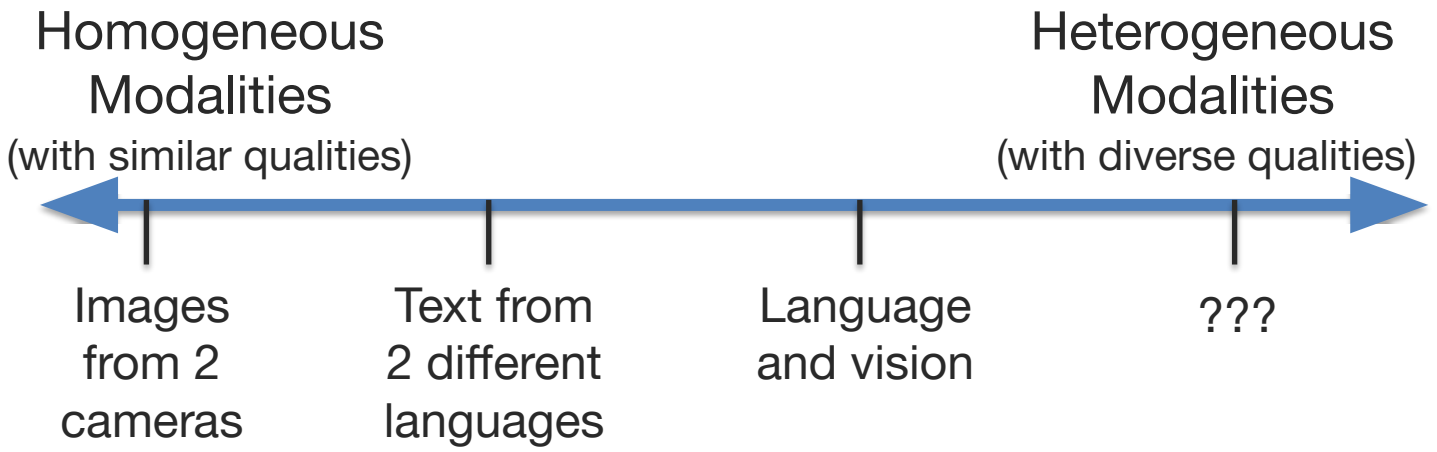
heterogeneous and interconnected data

Heterogeneous Modalities

Information present in different modalities will often show diverse qualities, structures and representations.



Examples:



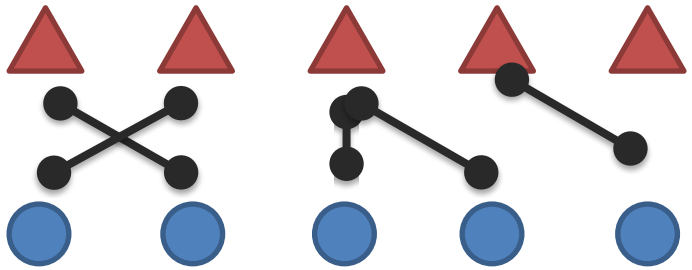
Abstract modalities are more likely to be homogeneous

What are the main Dimension of Multimodal Heterogeneity?

Interconnected Modalities

Modality A

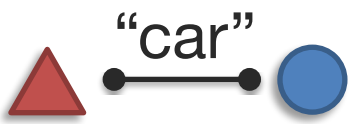
Modality B



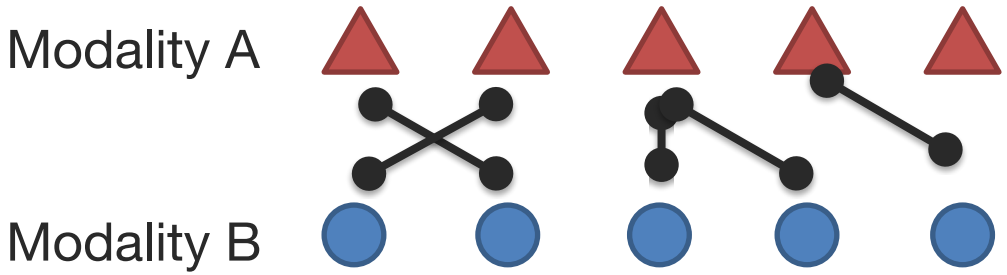
① Connections

Which elements are connected and why?

Statistical ⟵ ⟶ **Semantic**

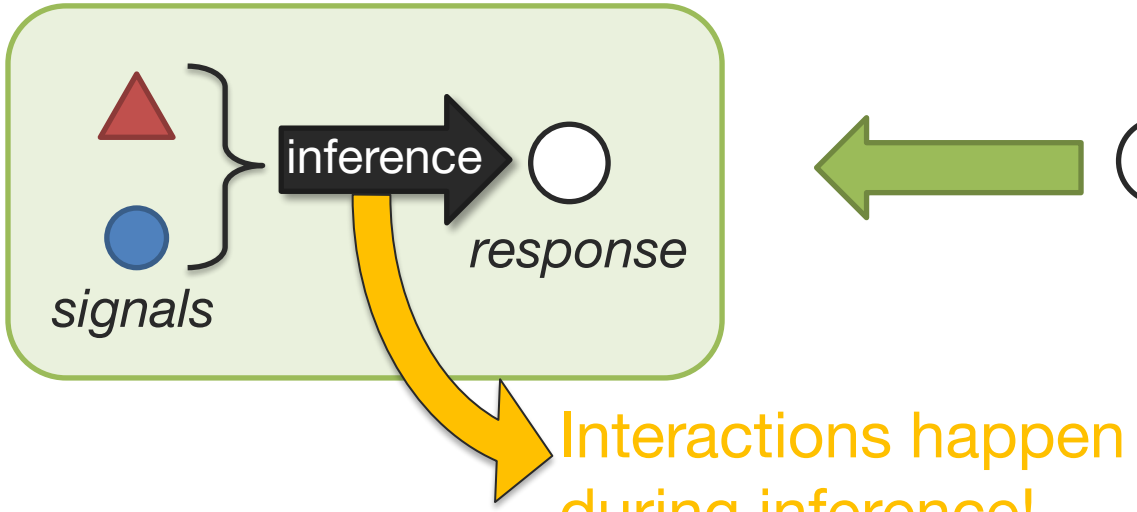
Association	Dependency	Correspondence	Relationship
			
e.g., correlation, co-occurrence	e.g., temporal contingency	e.g., grounding	e.g., knowledgebase

Interconnected Modalities



① Connections

Which elements should be connected and why?



② Cross-modal interactions

How are connected elements interacting during inference?

What are the Dimensions of Cross-Modal Interactions?

Multimodal Technical Challenges – Surveys, Tutorials and Courses

2016

Multimodal Machine Learning: A Survey and Taxonomy

Tadas Baltrusaitis, Chaitanya Ahuja and Louis-Philippe Morency

(Arxiv 2017, IEEE TPAMI journal, February 2019)

https://arxiv.org/abs/1705.09406

Tutorials: CVPR 2016, ACL 2016, ICMI 2016, …

Graduate-level courses:

Multimodal Machine learning (11th edition)

https://cmu-multicomp-lab.github.io/mmml-course/fall2020/

Advanced Topics in Multimodal Machine learning

https://cmu-multicomp-lab.github.io/adv-mmml-course/spring2022/

2022

Fundamentals of Multimodal ML: A Taxonomy & Open Challenges

Paul Liang, Amir Zadeh and Louis-Philippe Morency

- ☑ 6 core challenges
- ☑ 50+ taxonomic classes
- ☑ 600+ referenced papers

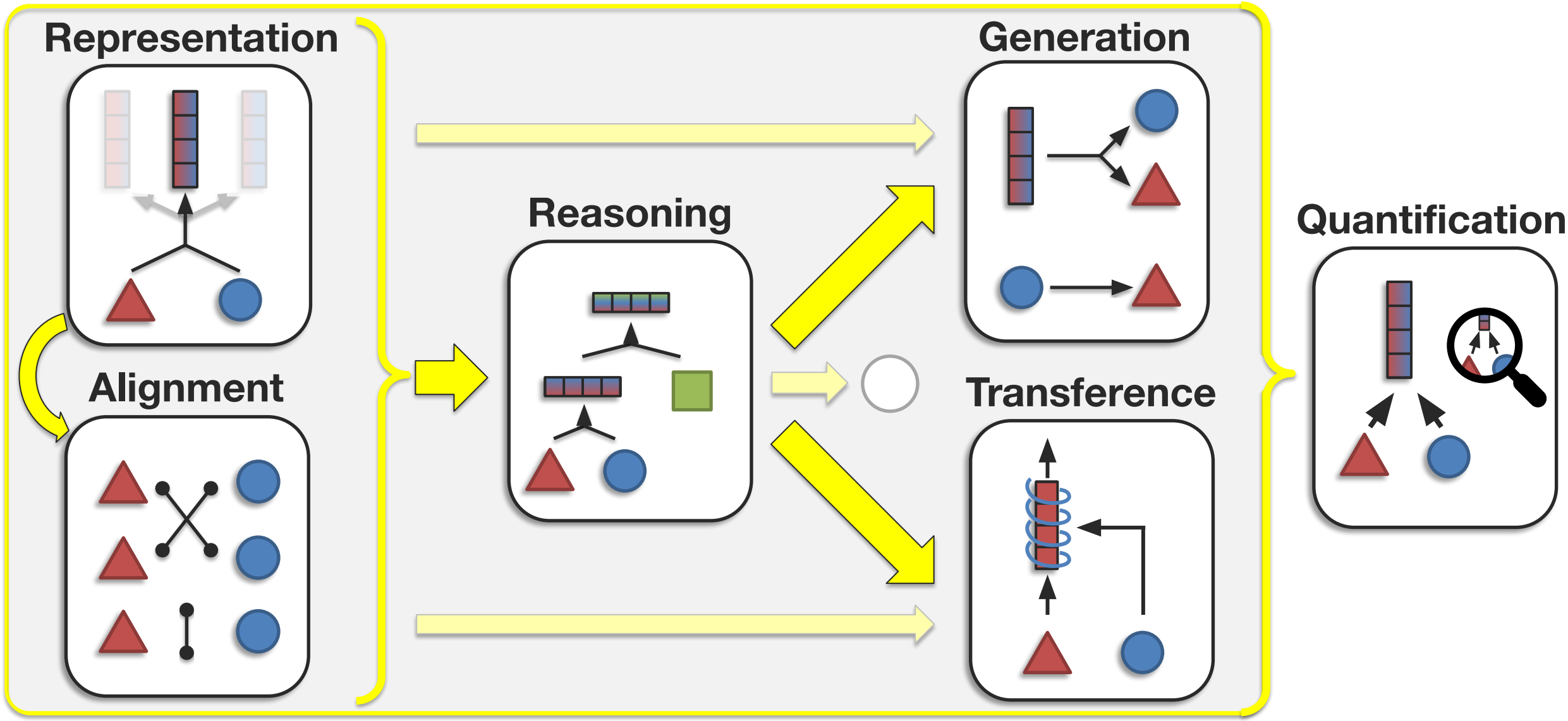
Tutorials: CVPR 2022, NAACL 2022, …

Updated graduate-level course:

Multimodal Machine learning (12th edition)

Fall 2022 semester

Core Multimodal Challenges



Any other Core Technical Challenges?

Course Syllabus

Learning Objectives

1. Study recent technical achievements in multimodal research
2. Improve critical and creative thinking skills
3. Understand future research challenges in multimodal
4. Explore new research ideas in multimodal learning

Two Versions: 6-credits and 12-credits

- 6-credit version:
  - Reading assignments
  - Small group discussions
  - Synopsis leads
- 12-credit version
  - Same 6-credit expectations + a high-quality research project:
    - Proposal with literature review
    - Midterm and final reports
    - Bi-weekly updates

Course Topics *(subject to change, based on student interests and course discussions)*

Week 1 (1/20): Introduction

Week 2 (1/27): Dimensions of heterogeneity

Week 3 (2/3): Connections and interactions – part 1

Week 4 (2/10): Connections and interactions - part 2

Week 5 (2/17): Modality utility and selection

Week 6 (2/24): Quantification and visualization

Week 7 (3/3): Empirical and theoretical frameworks

Week 8 (3/10): *No classes – Spring break*

Course Topics *(subject to change, based on student interests and course discussions)*

Week 9 (3/17): Brain and multimodal perception

Week 10 (3/24): Multimodal reasoning

Week 11 (3/31): Pretraining and scaling

Week 12 (4/7): *No classes – CMU Carnival*

Week 13 (4/14): Generalization and optimization

Week 14 (4/21): Open research questions

Week 15 (4/28): Report presentations

Reading Assignments

- Three main parts:
  - **Paper scouting:** Scout for extra papers, blog posts or other resources related to these question probes
  - **Reading notes:** Read the assigned papers and summarizing the main take-away points of each paper
    - Optional: if you have clarification questions about the papers
  - **Discussion points:** Reflect on the question probes related to the reading papers and prepare discussion points.
- 12 readings assignments, with usually 2 required papers and some suggested (but optional) papers

How Each Weekly Class will Happen

- Joint portion (about 15 mins)
  - Short presentation presenting the scouted papers and answering student questions about the required papers
- Separate discussion groups (about 1 hours)
  - Two groups of 8-10 students, one instructor per group
  - Round-table discussions: Discuss the research question probes. Each student is expected to actively participate in this discussion.
  - Two note-takers per discussion groups (alternating note-taking)

Discussion Roles

Reading leads (1 per discussion group, 2 total per week):

1. Short presentation (10-15 mins)
   a) Answer questions from other students
   b) Summarize scouted papers
2. Help with note-taking during discussions

Synopsis leads (1 per discussion group, 2 total per week):

3. Note taking during discussions
4. Prepare discussion synopsis
   a) Summarizing the main discussion points
   b) Overview schema, table or figure

What Weeks Would you Prefer to Lead?

Week 2 (1/27): **Heterogeneity**

Week 3 (2/3): **Connections & interactions 1**

Week 4 (2/10): **Connections & interactions 2**

Week 5 (2/17): **Modality utility & selection**

Week 6 (2/24): **Quantification & visualization**

Week 7 (3/3): **Empirical & theoretical frameworks**

Week 9 (3/17): **Brain & multimodal**

Week 10 (3/24): **Multimodal reasoning**

Week 11 (3/31): **Pretraining & scaling**

Week 13 (4/14): **Generalization & optimization**

Week 14 (4/21): **Open challenges**

Grading Scheme for 6-credit Version

- Reading assignments 40%
  - 4 points per assignment
  - Top 10 scores kept for final grade
- Participation and discussions 40%
  - 4 points per discussion session
  - Top 10 scores kept for final grade
- Reading and synopsis leads 20%
  - Reading leads: 5 points for each presentation (including note-keeping)
    - Top 2 scores are kept for final grade
  - Synopsis leads: 5 points for each synopsis (including note-keeping)
    - Top 2 scores are kept for final grade

Research Course Project (12-credit version)

- ✔ Similar in spirit to a 6-credit independent study project
- ✔ Project teams of 2 or 3 students
- ✔ Final report should be like a research paper
- ✔ Expected to explore new research ideas
- ✔ Regular meetings with instructors

Course Project Topics

- Quantifying heterogeneity
  - Modality-general model with modality-specific components that are automatically activated depending on heterogeneity?
- Quantifying and visualizing modality interactions
  - Formal measures of redundancy, uniqueness, synergy; statistical feature interactions
- Modality tradeoffs & dynamic modality selection
  - Connections to feature selection; benefits and risks of modalities
- Empirical & theoretical frameworks to explain multimodal phenomenon
  - Modality benefits, optimization challenges, modality collapse, multimodal pretraining
- Multimodal with non-deep-learning effective modalities (e.g., tabular, time-series)
- Efficiency, compression, sparse multimodal models
- Gesture generation, conditioning LLM on other modalities, see more on piazza

Bi-weekly Project Meetings and Updates

- Required meetings on a bi-weekly basis
  - About 20 minutes per meeting, usually on Thursdays or Fridays
  - Primary mentor (Paul or LP) for each team
- Bi-weekly written updates
  - Either Google Slides (preferred) or Google Docs
  - Due Tuesdays at 9pm before the meeting (due Monday 9pm for reports)
  - Some expectations for each bi-weekly update (see next slide)
- Alternate weeks: optional meetings with either mentor
  - Sign-up website for meetings with either LP or Paul
  - No written update required, but suggested

Schedule for Bi-Weekly Written Updates and Reports

- Week 3: Pre-proposal details with initial literature review
- Week 5: **Proposal report:** literature review + baseline selection
- Week 7: Results with baselines and initial implementation of idea
- *Week 8: Spring break (no meetings, no work, relax 🙂)*
- Week 10: **Midterm report:** first complete round of results for idea
- Week 12: Updated results for research idea
- Week 14: Error analysis, ablations and visualizations
- Week 15:(Friday 4/28) Poster presentations
  - Tuesday 5/2 at 9pm**: final report**

Course Project Timeline

- **Project preferences** (Due Tuesday 1/24 at 9pm ET) –share your interests about research projects, to help with team matching.
- **Pre-proposal** (Due Tuesday 1/31 at 9pm ET) – You should have selected your teammates, have ideas about your dataset and task.
- **Proposal and Literature Review** (Due Monday 2/13 at 9pm ET) –Research ideas, review of relevant papers and initial results
- **Midterm report** (Due Monday 3/20 at 9pm ET) – Intermediate report documenting the updated results exploring your research ideas.
- **Final report** (Due Monday 5/2 at 9pm ET) – Final report describing explored research ideas, with results, analysis and discussion.

Grading Scheme for 12-credit Version

- Grading breakdown of the 6-unit version will be scaled to 50%.

- The second 50% comes from the course project:
  - Proposal report 10%
  - Midterm report 20%
  - Final report 30%
  - Final presentation 10%
  - Bi-weekly written updates 30%
    - 10 points per update, top 3 scores kept for final grade (out of 4 updates)

Absences and Late Submissions

- Lectures are not recorded, students expected to attend live
  - If you plan to miss more than one lecture this semester, let us know as soon as possible.
- Reading assignment wildcards (3 per students)
  - 24-hours extension, max 1 per week
- Project assignment wildcards (2 per teams)
  - 24-hours extension, can be used together

Course Websites

- Piazza
  - For course announcements and assignments

  https://piazza.com/cmu/spring2023/11877/info
- CMU Canvas
  - For assignment submissions and grading

  https://canvas.cmu.edu/
- Course website
  - A general public version of the course information
    - Discussion synopsis will be posted here

  https://cmu-multicomp-lab.github.io/adv-mmml-course/spring2023/

Assignments for This Coming Week

Week 2 reading assignment (Due Wednesday 1/25 at 9pm ET)

- Detailed instructions will be posted on Piazza
  - Required paper: Geometric deep learning, a unified paradigm to reason about structure, invariance, properties, and inductive biases in each modality.
  - Suggested papers: Useful dimensions of heterogeneity in domain adaptation, transfer learning, multitask learning, quantifying dimensions of heterogeneity.

For students taking the 12-credit version:

- Project preference form (Due this Tuesday 1/24 at 9pm ET)
  - To help with team matching
  - Google Form link is also available Piazza

    https://forms.gle/QzJuVjzGDQwxgrH8A