Language Technologies Institute

Carnegie Mellon University

Multimodal Machine Learning

Lecture 1.1: Introduction

Louis-Philippe Morency

** Fall 2021, 2022 and 2023 co-lecturer: Paul Liang. Original course co-developed with Tadas Baltrusaitis. Spring 2021 and 2022 editions taught by Yonatan Bisk. Spring 2023 edition taught by Yonatan and Daniel Fried*

Your teaching team This Semester (11-777, Fall 2023)

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

Paul Liang
pliang@cs.cmu.edu
Co-lecturer

Syeda Akter
sakter@andrew.cmu.edu
TA

Mehul Agarwal
mehula@andrew.cmu.edu
TA

Aditya Rathod
arathod@andrew.cmu.edu
TA

Soham Dinesh Tiwari
sohamdit@andrew.cmu.edu
TA

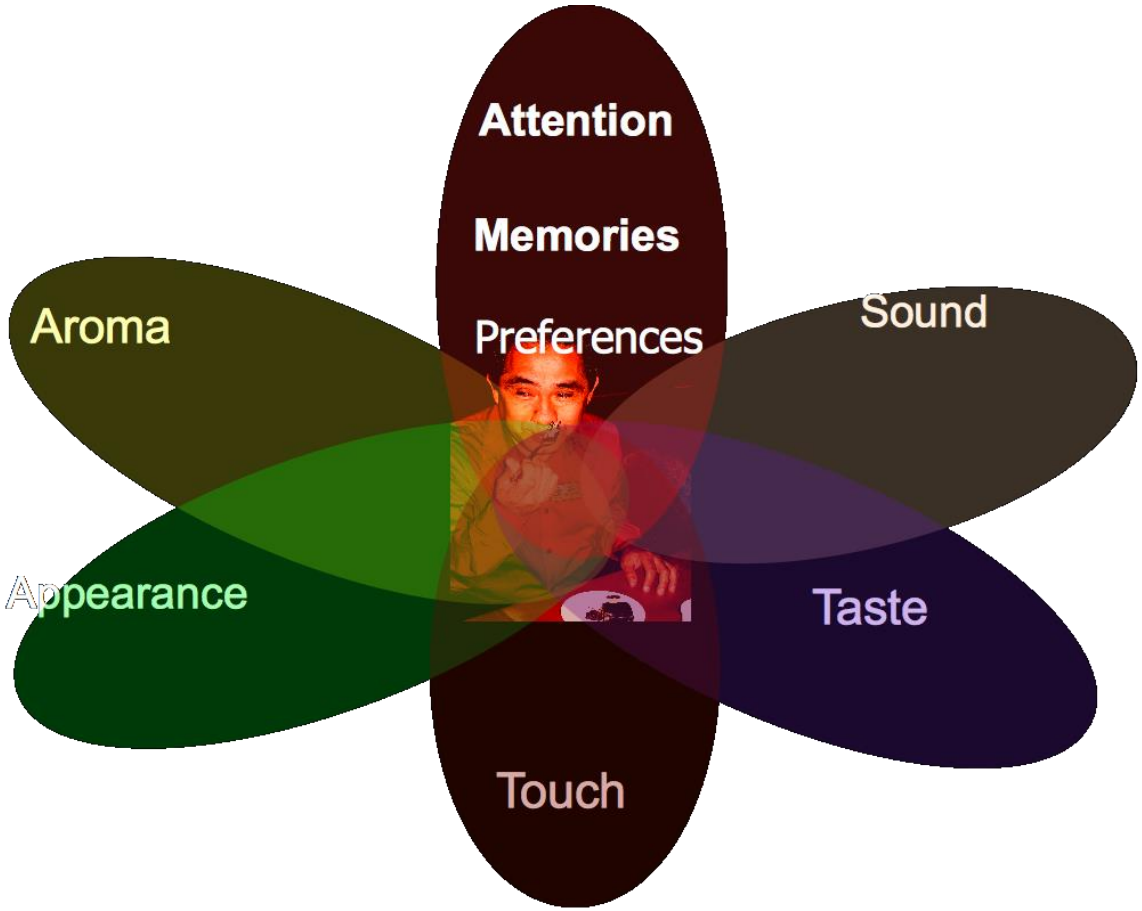
Haofei Yu
haofeiy@andrew.cmu.edu
TA

Lecture Objectives

- What is Multimodal?
  - Research-oriented definition
  - Dimensions of modality heterogeneity
  - Modality connections and interactions
- Core technical and conceptual challenges
  - Representation, alignment, reasoning, generation, transference and quantification
- Course syllabus

What is Multimodal?

What is Multimodal?



Sensory Modalities

Multimodal Behaviors and Signals

Language

- **Lexicon**
  - Words
- **Syntax**
  - Part-of-speech
  - Dependencies
- **Pragmatics**
  - Discourse acts

Acoustic

- **Prosody**
  - Intonation
  - Voice quality
- **Vocal expressions**
  - Laughter, moans

Visual

- **Gestures**
  - Head gestures
  - Eye gestures
  - Arm gestures
- **Body language**
  - Body posture
  - Proxemics
- **Eye contact**
  - Head gaze
  - Eye gaze
- **Facial expressions**
  - FACS action units
  - Smile, frowning

Touch

- **Haptics**
- **Motion**

Physiological

- **Skin conductance**
- **Electrocardiogram**

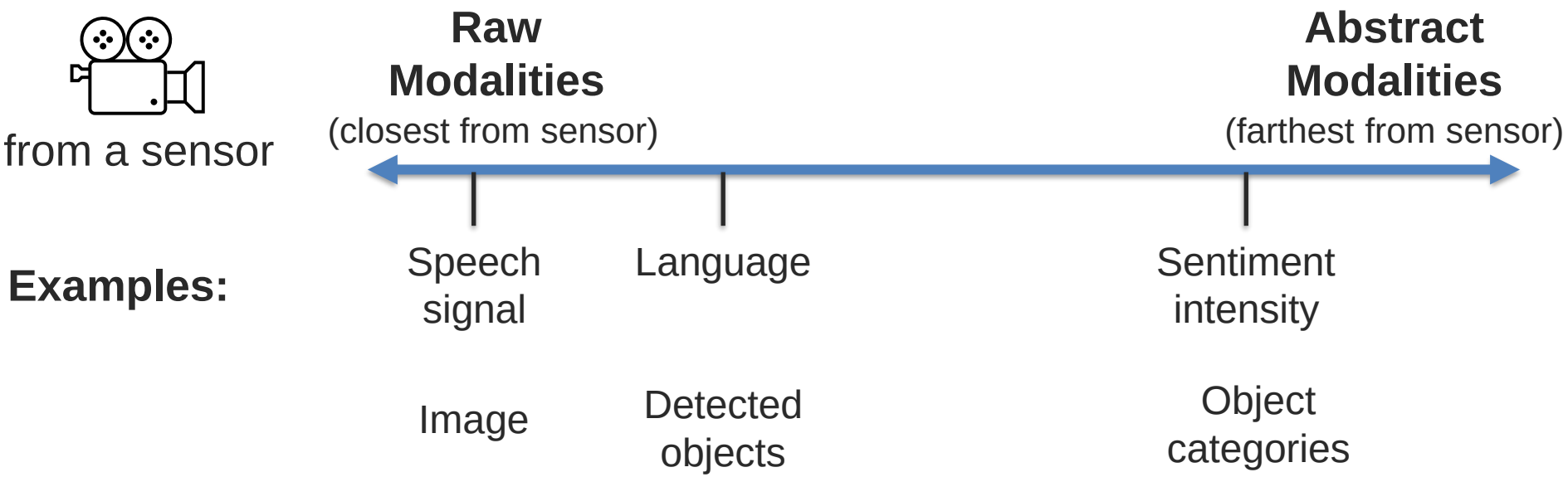
Mobile

- **GPS location**
- **Accelerometer**
- **Light sensors**

What is a Modality?

Modality

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



What is Multimodal?

A dictionary definition…

Multimodal: with multiple modalities

A research-oriented definition…

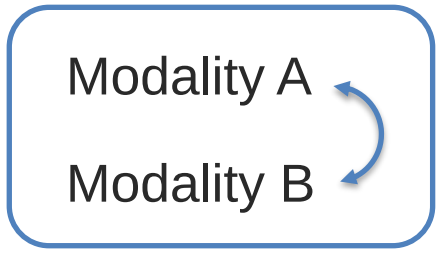
***Multimodal* is the scientific study of**

heterogeneous and interconnected data

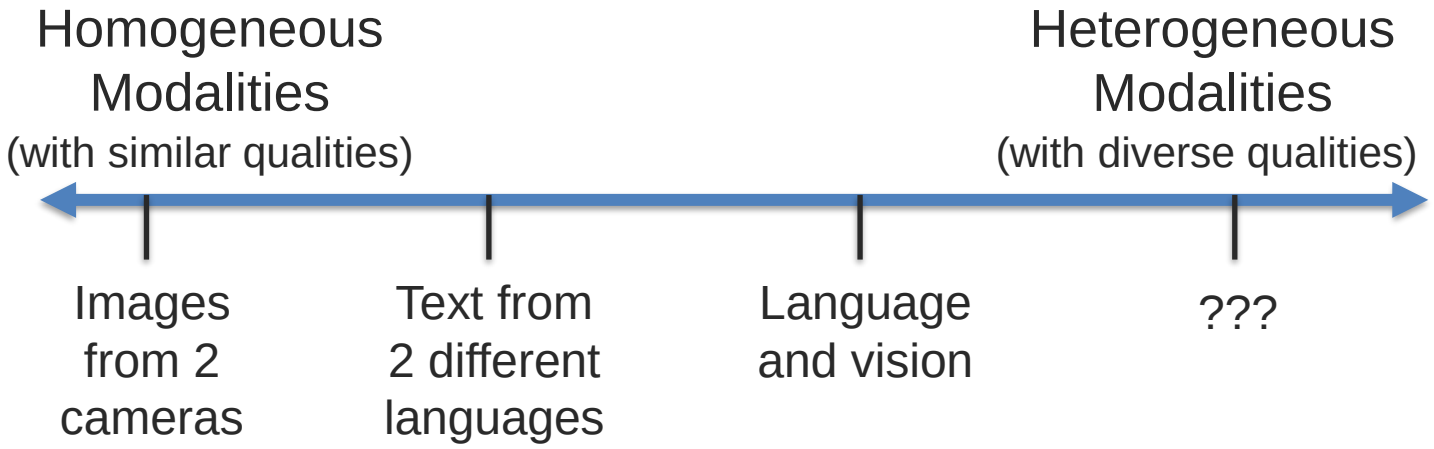
Connected + Interacting

Heterogeneous Modalities

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



Examples:



Abstract modalities are more likely to be homogeneous

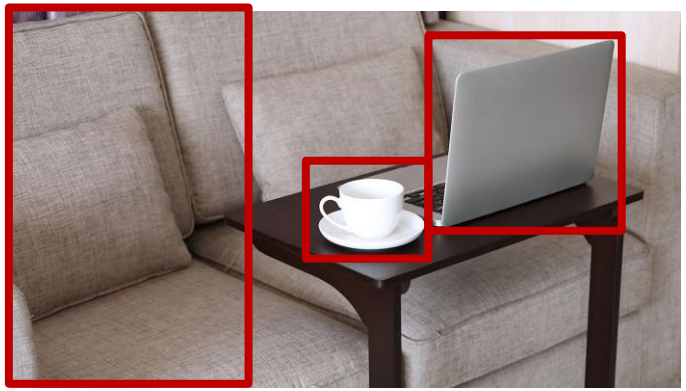
Dimensions of Heterogeneity

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

A teacup on the right of a laptop in a clean room.

Dimensions of Heterogeneity

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



*A **teacup** on the **right** of a **laptop** in a **clean room**.*

1. **Element representations:** discrete, continuous, granularity

▲ {  ,  ,  }

● *{teacup, right, laptop, clean, room}*

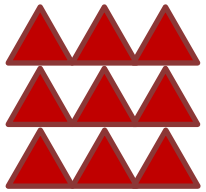
Dimensions of Heterogeneity

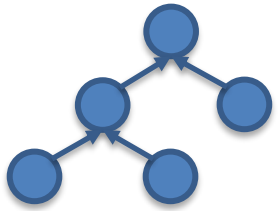
Modality A ⟷ Modality B

1. **Element representations:**
   Discrete, continuous, granularity

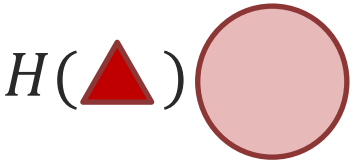
2. **Element distributions:**
   Density, frequency

3. **Structure:**
   Temporal, spatial, latent, explicit





4. **Information:**
   Abstraction, entropy

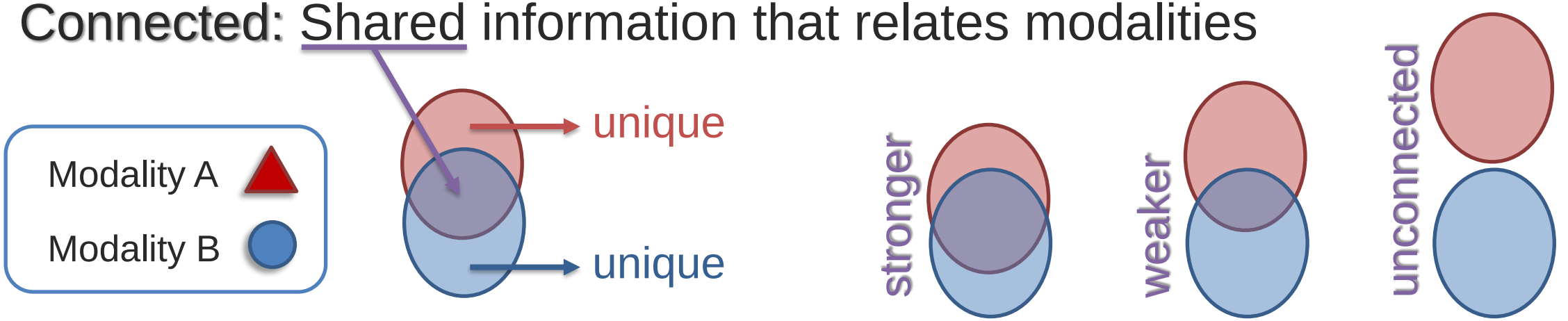


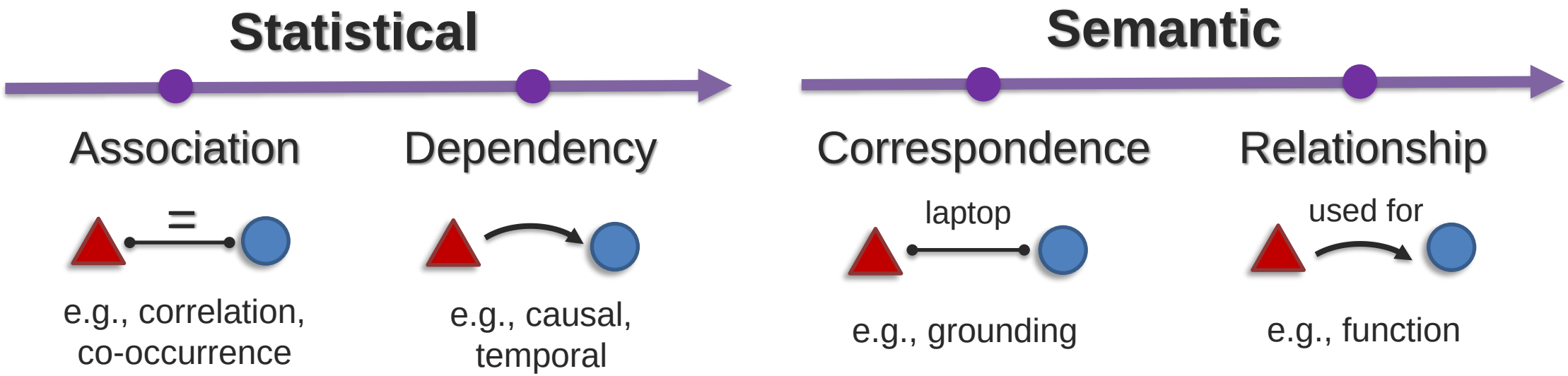
5. **Noise:**
   Uncertainty, noise, missing data

6. **Relevance:**
   Task, context dependence

Connected Modalities

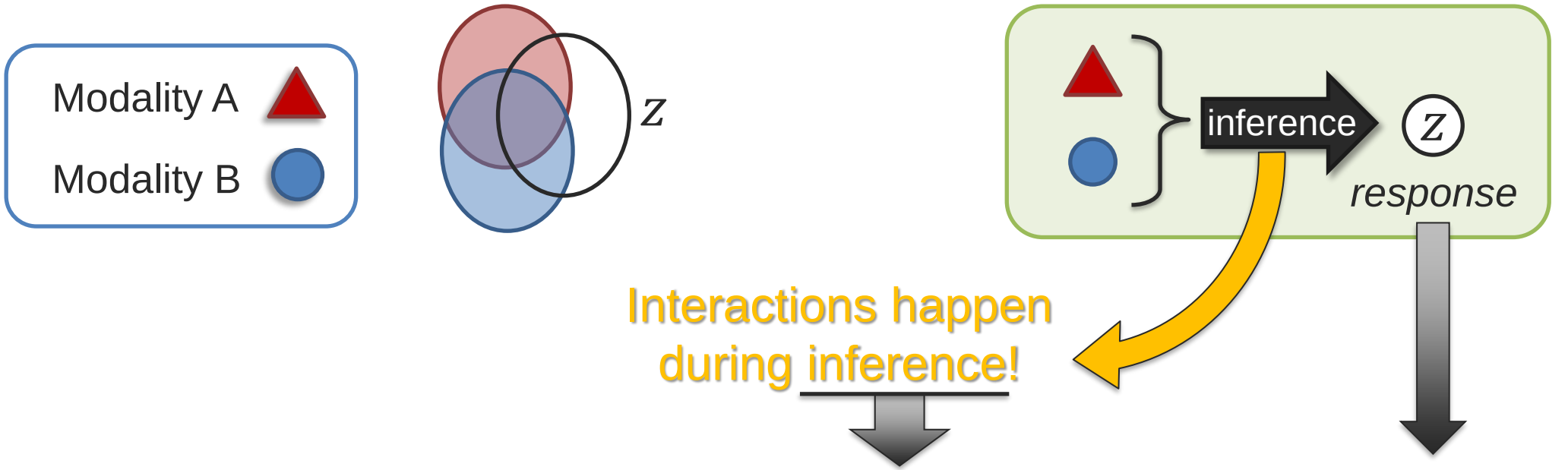
Connected: Shared information that relates modalities





Interacting Modalities

Interacting: process affecting each modality, creating new response



"Inference" examples:

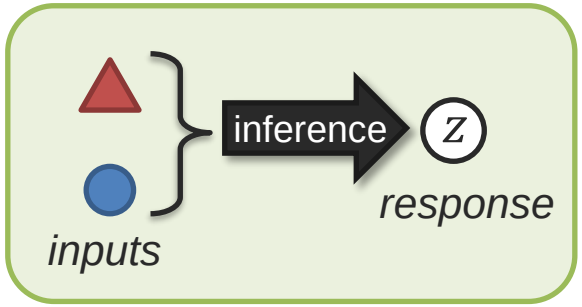
- Representation fusion
- Prediction task
- Modality translation

representation

$\hat{y}$ prediction

modality C

Taxonomy of Interaction Responses – A Behavioral Science View



Multimodal Communication

	signal	response	signal	response	
Redundancy	a → □		a+b → □		Equivalence
	b → □		a+b → ▭		Enhancement
Nonredundancy	a → □		a+b → □ and ○		Independence
	b → ○		a+b → □		Dominance
			a+b → ▭ (or □)		Modulation
			a+b → △		Emergence

Partan and Marler (2005). *Issues in the classification of multimodal communication signals*. American Naturalist, 166(2)

What is Multimodal?

Multimodal **is the scientific study of**

heterogeneous and interconnected data ☺

Multimodal Machine Learning

What is Multimodal Machine Learning?

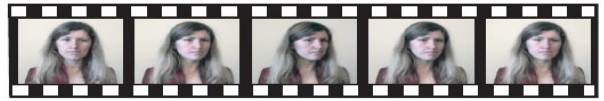
Multimodal Machine Learning (ML) is the study of computer algorithms that learn and improve through the use and experience of data from multiple modalities

Multimodal Artificial Intelligence (AI) studies computer agents able to demonstrate intelligence capabilities such as understanding, reasoning and planning, through multimodal experiences, and data

Multimodal AI is a superset of **Multimodal ML**

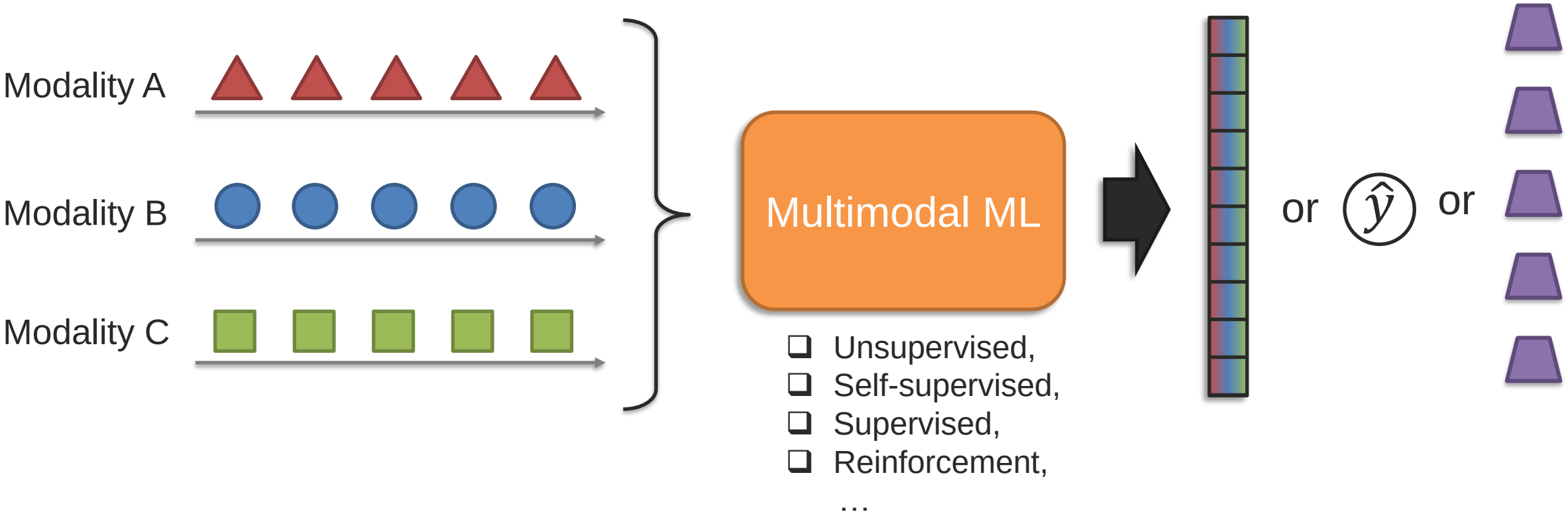
Multimodal Machine Learning

Language I really like this tutorial

Vision 

Acoustic 

Multimodal Machine Learning



- Unsupervised,
- Self-supervised,
- Supervised,
- Reinforcement,

…

Multimodal Machine Learning

*What are the **core multimodal technical challenges,***

understudied in conventional machine learning?

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

Foundations and Recent Trends in Multimodal Machine Learning

Paul Liang, Amir Zadeh and Louis-Philippe Morency

☑ 6 core challenges
☑ 50+ taxonomic classes
☑ 700+ referenced papers

https://arxiv.org/abs/2209.03430

Tutorials: ICML 2023, CVPR 2022, NAACL 2022

Updated graduate-level course:

Multimodal Machine learning (12th edition)

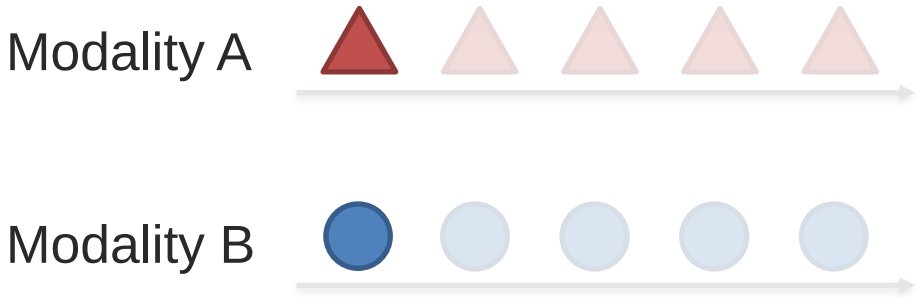
https://cmu-multicomp-lab.github.io/mmml-course/fall2022/

Challenge 1: Representation

Definition: Learning representations that reflect cross-modal interactions between individual elements, across different modalities

➡ This is a core building block for most multimodal modeling problems!

Individual elements:



It can be seen as a "local" representation

or

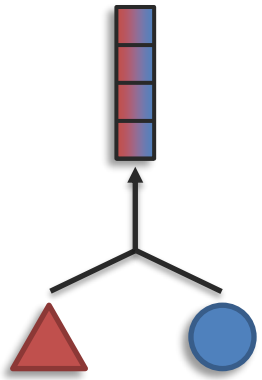
representation using holistic features

Challenge 1: Representation

Definition: Learning representations that reflect cross-modal interactions between individual elements, across different modalities

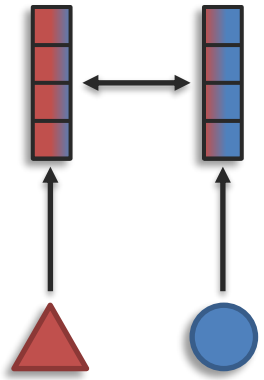
Sub-challenges:

Fusion



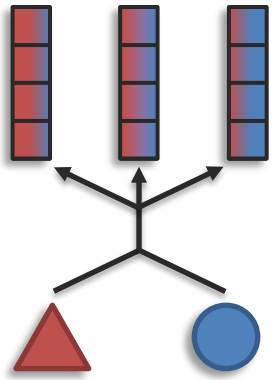
modalities > # representations

Coordination



modalities = # representations

Fission



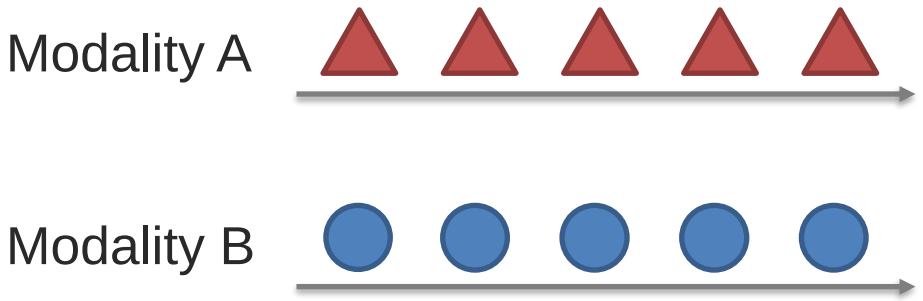
modalities < # representations

Challenge 2: Alignment

Definition: Identifying and modeling cross-modal connections between all elements of multiple modalities, building from the data structure

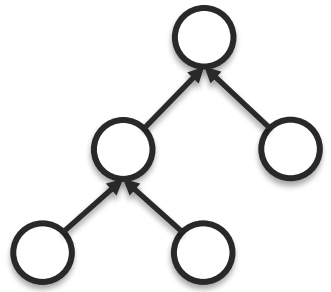
➡ Most modalities have internal structure with multiple elements

Elements with temporal structure:



Other structured examples:

Spatial



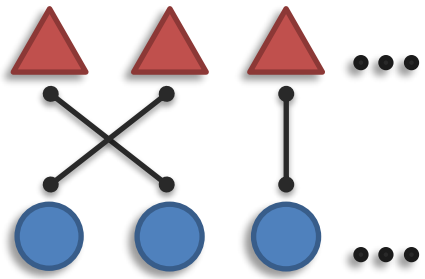
Hierarchical

Challenge 2: Alignment

Definition: Identifying and modeling cross-modal connections between all elements of multiple modalities, building from the data structure

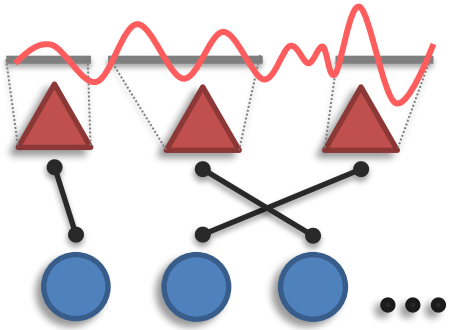
Sub-challenges:

Discrete Alignment



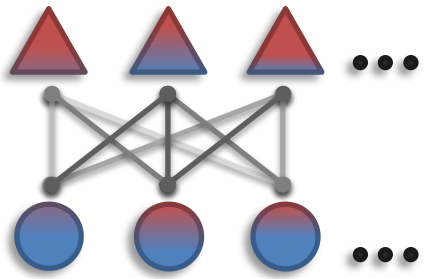
Discrete elements and connections

Continuous Alignment



Segmentation and continuous warping

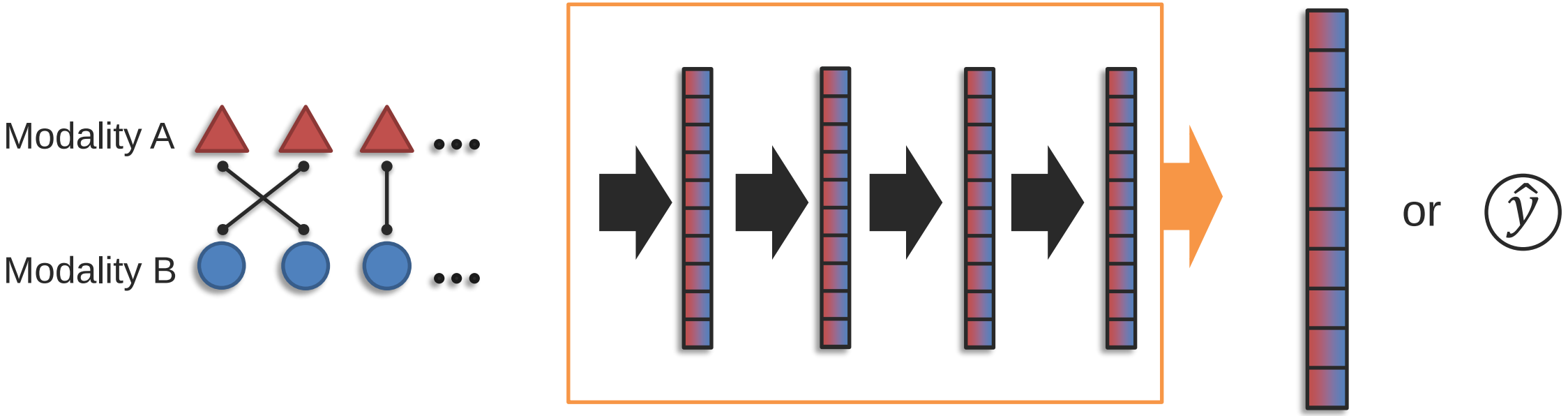
Contextualized Representation



Alignment + representation

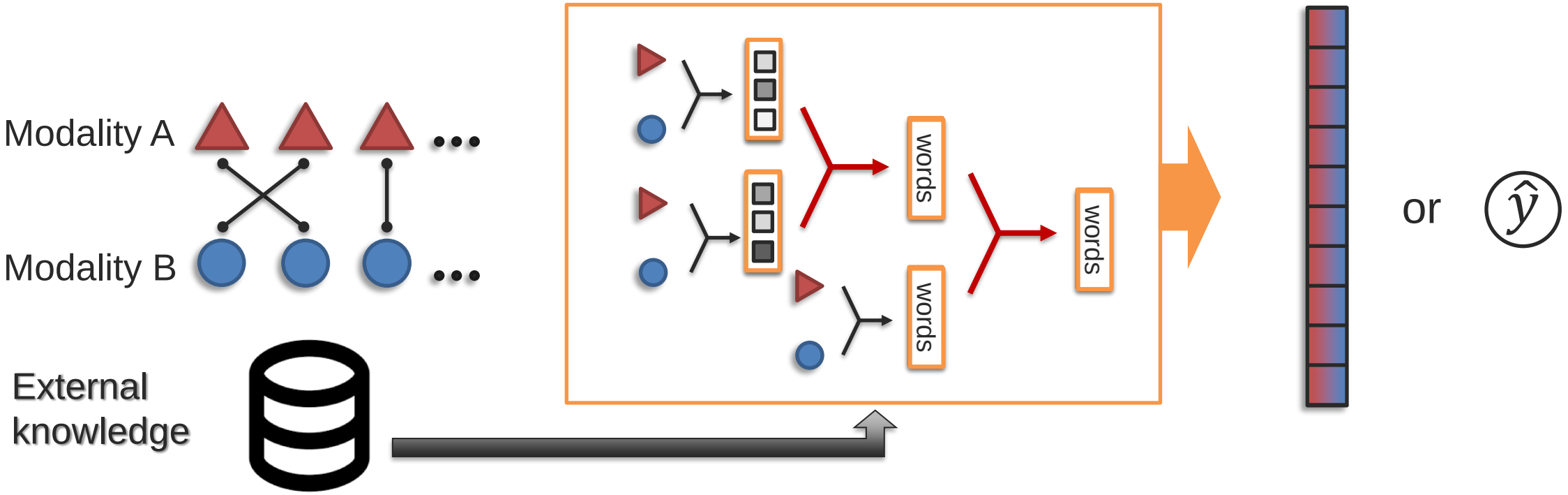
Challenge 3: Reasoning

Definition: Combining knowledge, usually through multiple inferential steps, exploiting multimodal alignment and problem structure



Challenge 3: Reasoning

Definition: Combining knowledge, usually through multiple inferential steps, exploiting multimodal alignment and problem structure

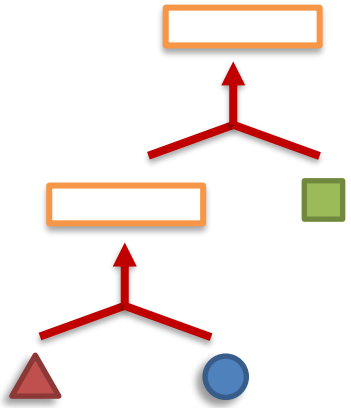


Challenge 3: Reasoning

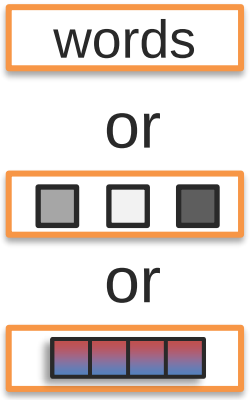
Definition: Combining knowledge, usually through multiple inferential steps, exploiting multimodal alignment and problem structure

Sub-challenges:

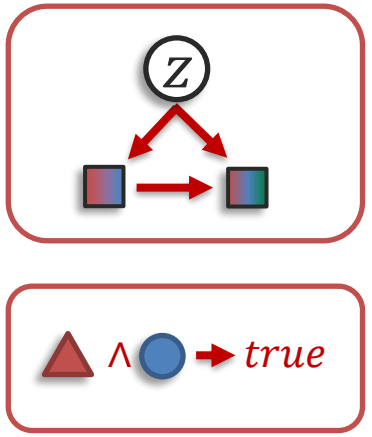
Structure Modeling



Intermediate concepts



Inference Paradigm

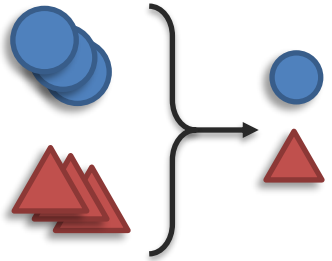
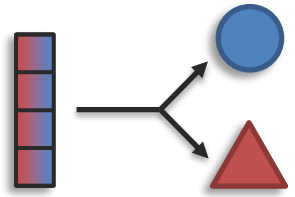


External Knowledge

Challenge 4: Generation

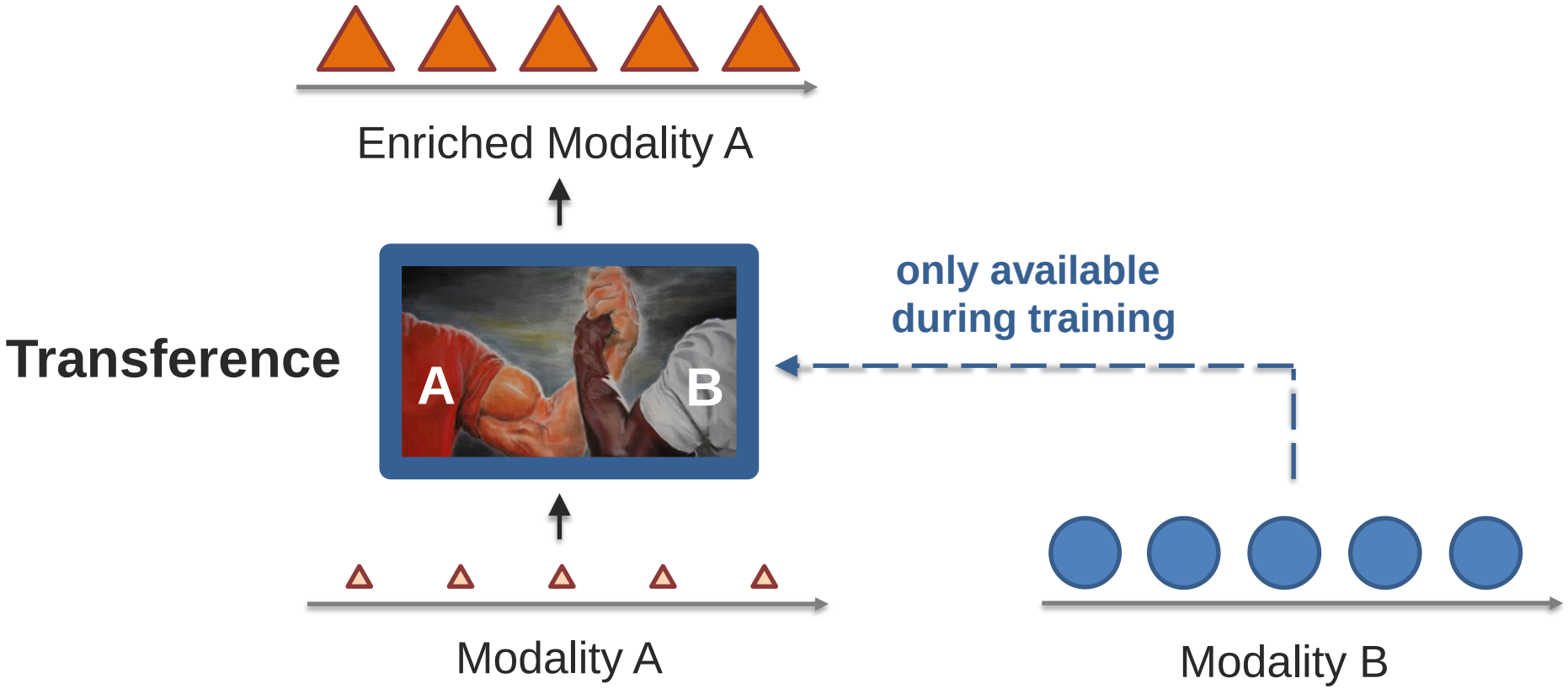
Definition: Learning a generative process to produce raw modalities that reflects cross-modal interactions, structure and coherence

Sub-challenges:

	Summarization	Translation	Creation
			
Information: (content)	Reduction	Maintenance	Expansion
			

Challenge 5: Transference

Definition: Transfer knowledge between modalities, usually to help the target modality which may be noisy or with limited resources

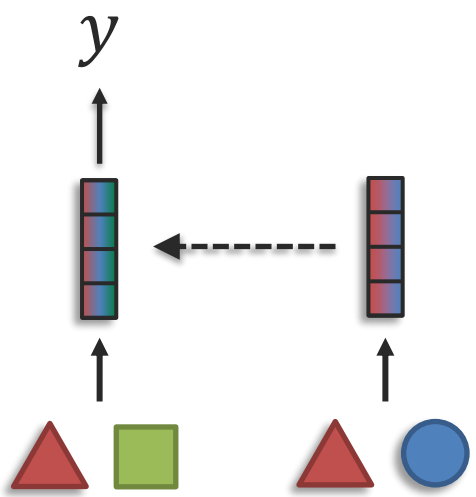


Challenge 5: Transference

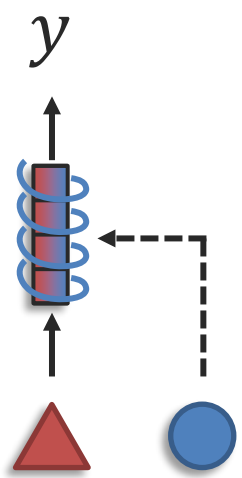
Definition: Transfer knowledge between modalities, usually to help the target modality which may be noisy or with limited resources

Sub-challenges:

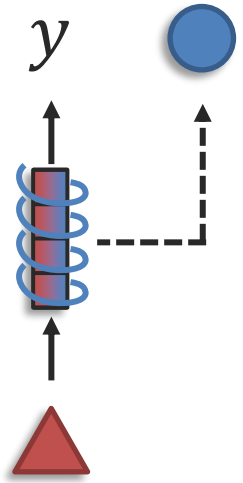
Transfer



Co-learning
via representation



Co-learning
via generation

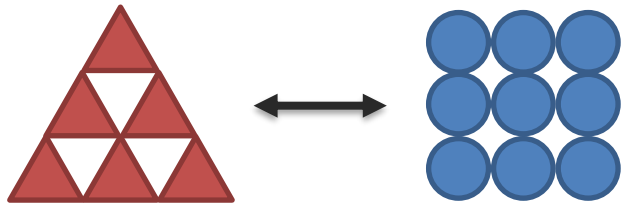


Challenge 6: Quantification

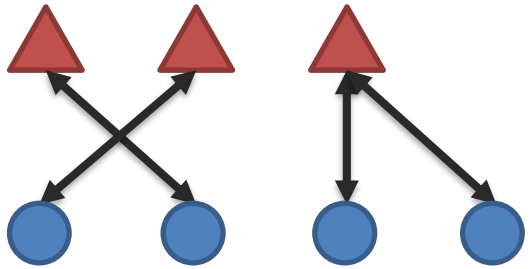
Definition: Empirical and theoretical study to better understand heterogeneity, cross-modal interactions and the multimodal learning process

Sub-challenges:

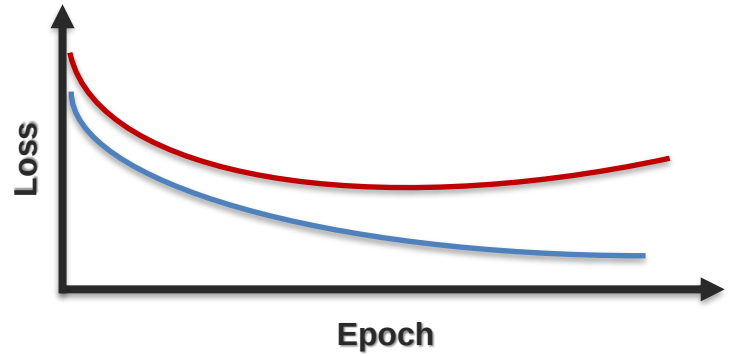
Heterogeneity



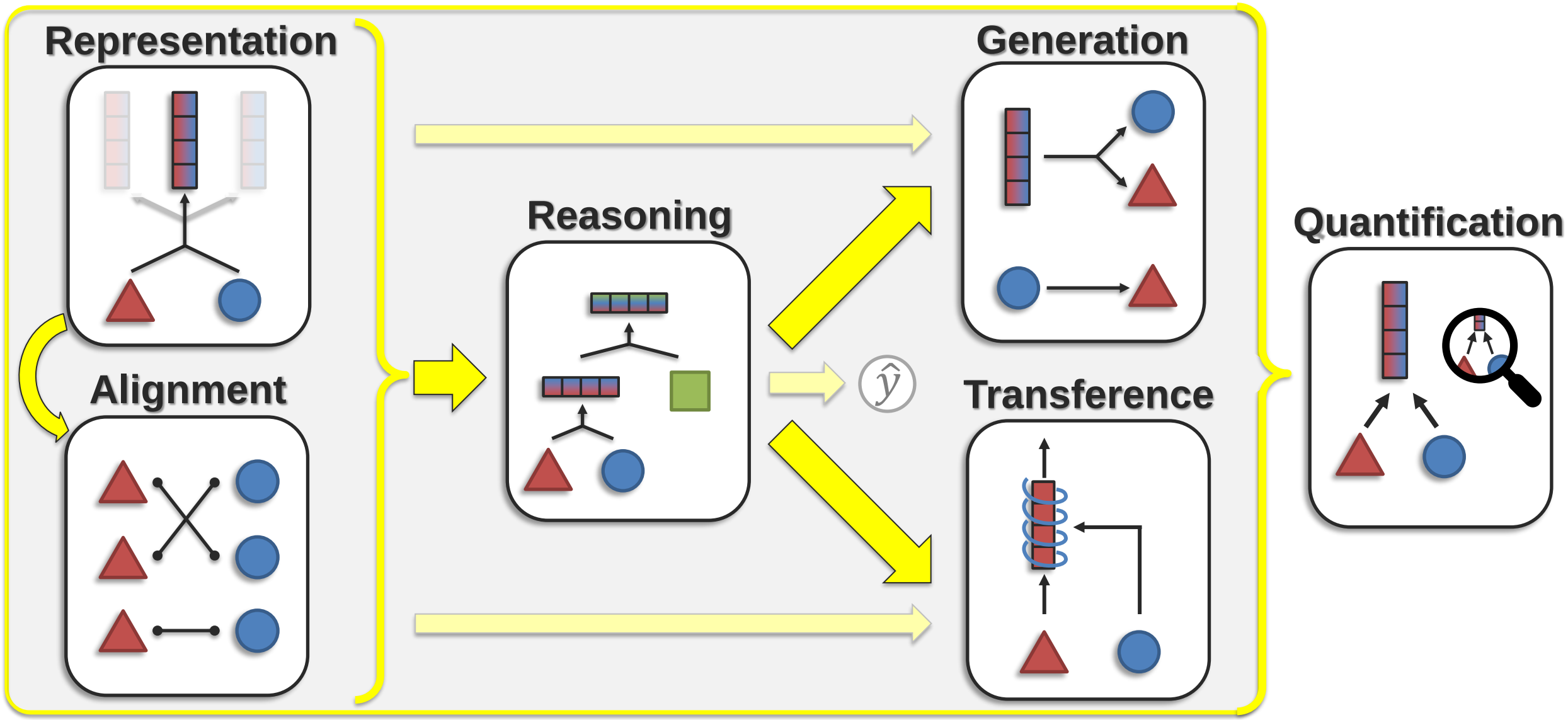
Interactions



Learning



Core Multimodal Challenges



Lecture Schedule

Classes	Tuesday Lectures	Thursday Lectures
Week 1 8/29 & 8/31	**Course introduction** • Multimodal core challenges • Course syllabus	**Multimodal applications and datasets** • Research tasks and datasets • Team projects
Week 2 9/5 & 9/7 Read due: 9/9	**Unimodal representations** • Dimensions of heterogeneity • Visual representations	**Unimodal representations** • Language representations • Signals, graphs and other modalities
Week 3 9/12 & 9/14 Read due: 9/16 Proj. Due: 9/13	**Multimodal representations** • Cross-modal interactions • Multimodal fusion	**Multimodal representations** • Coordinated representations • Multimodal fission
Week 4 9/19 & 9/21 *Proj. due: 9/24*	**Multimodal alignment and grounding** • Explicit alignment • Multimodal grounding	**Alignment and representations** • Self-attention transformer models • Masking and self-supervised learning
Week 5 9/26 & 9/28 Read due: 9/30	**Multimodal transformers** • Multimodal transformers • Video and graph representations	**Multimodal Reasoning** • Structured and hierarchical models • Memory models
Week 6 10/3 & 10/5 *Proj. due: 10/8*	***Project hours***	***Multimodal language grounding*** • Grounded semantics and pragmatics

Lecture Schedule

Classes	Tuesday Lectures	Thursday Lectures
Week 7 10/10 & 10/12 Read due: 10/14	**Multimodal interaction** • Reinforcement learning • Discrete structure learning	**Multimodal inference** • Logical and causal inference • External knowledge
Week 8 10/17 & 10/19	***Fall Break – No lectures***	
Week 9 10/24 & 10/26 *Proj. due: 10/29*	**Multimodal generation** • Translation, summarization, creation • Generative models: VAEs	**New generative models** • GANs and diffusion models • Model evaluation and ethics
Week 10 10/31 & 11/2	***Project presentations (midterm)***	***Project presentations (midterm)***
Week 11 11/7 & 11/9 Read due: 11/12	***Democracy Day – No Class –***	**Transference** • Modality transfer and co-learning • Self-training and multitask learning
Week 12 11/14 & 11/16 Read due: 11/21	**Quantification** • Heterogeneity and interactions • Biases and fairness	**New research directions** • Recent research in multimodal ML

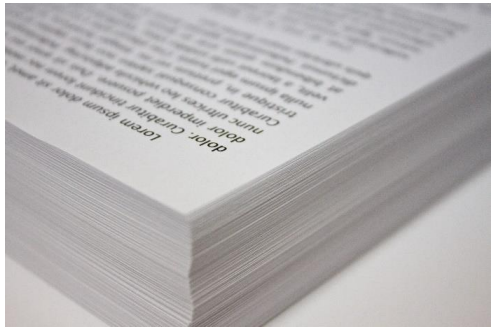
Lecture Schedule

Classes	Tuesday Lectures	Thursday Lectures
Week 13 11/21 & 11/23	***Thanksgiving Week – No Class –***	
Week 14 11/28 & 11/30	**Guest lecture**	**Guest lecture**
Week 15 12/5 & 12/7 *Proj. due: 12/10*	***Project presentations (final)***	***Project presentations (final)***

Course Syllabus

Three Course Learning Paradigms

Course lecture participation
(16% of your grade)



Reading assignments
(12% of your grade)

$$
\begin{aligned}
i_t &= \sigma\left(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i\right) \\
f_t &= \sigma\left(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f\right) \\
c_t &= f_t c_{t-1} + i_t \tanh\left(W_{xc}x_t + W_{hc}h_{t-1} + b_c\right) \\
o_t &= \sigma\left(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o\right) \\
h_t &= o_t \tanh(c_t)
\end{aligned}
$$

Course project assignments
(72% of your grade)

Course Recommendations and Requirements

1. Ready to read about 6 papers this semester !
   - Curated list of research papers for the 6 reading assignments
   - Summarize one paper and contrast it with other papers
2. Already taken a machine learning course
   - Strongly recommended for students to have taken an introduction machine learning course
   - 10-401, 10-601, 10-701, 11-663, 11-441, 11-641 or 11-741
3. Motivated to produce a high-quality course project
   - Projects are designed to enhance state-of-the-art algorithms
   - Four project assignments, to help scaffold the project tasks

Course Project Guidelines

- Dataset should have at least two modalities:
  - Natural language and visual/images
- Teams of 3, 4 or 5 students
- The project should explore algorithmic novelty
- Possible venues for your final report:
  - NAACL 2024, ACL 2024, IJCAI 2024, ICML 2024, ICMI 2024
- We will discuss on Thursday about project ideas
- GPU resources available:
  - Amazon AWS and Google Cloud Platform

Course Project Timeline

$$
\begin{aligned}
i_t &= \sigma\left(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i\right)\\
f_t &= \sigma\left(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f\right)\\
c_t &= f_t c_{t-1} + i_t \tanh\left(W_{xc}x_t + W_{hc}h_{t-1} + b_c\right)\\
o_t &= \sigma\left(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o\right)\\
h_t &= o_t \tanh(c_t)
\end{aligned}
$$

Pre-proposal *(due Wednesday Sept. 13)*

- Define your dataset, research task and teammates

First project assignment *(due Sunday Sept. 24)*

- Study related work to your selected research topic

Second project assignment *(due Sunday Oct 8)*

- Experiment with unimodal representations

Midterm project assignment (due Sunday Oct 29)

- Implement and evaluate state-of-the-art model(s)

Final project assignment (due Sunday Dec. 10)

- Implement and evaluate new research ideas

# Equal Contribution by All Teammates!

- Each team will be required to create a GitHub repository which will be accessible by TAs
- Each report should include a description of the task from each teammate
- Please let us know soon if you have concerns about the participation levels of your teammates

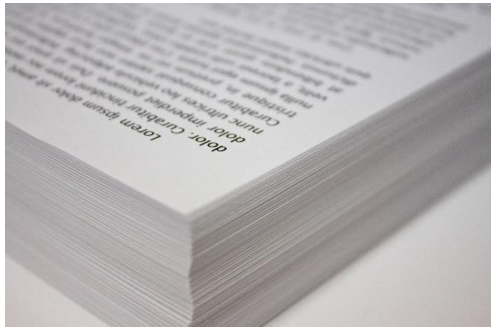
Process for Selecting your Course Project

- **Thursday 8/31:** Lecture describing available multimodal datasets and research topics
- **Tuesday 9/5:** Let us know your dataset preferences for the course project
- **Thursday 9/7:** During the later part of the lecture, we will have an interactive period to help with team formation. More details to come
- **Wednesday 9/13:** Pre-proposals are due. You should have selected your teammates, dataset and task

Project Preferences – Due Tuesday 9/6

- Post your project preferences:
  - List of your ranked preferred projects
    - Use alphanumeric code of each dataset
    - Detailed dataset list in the "Lecture1.2-datasets“ slides
  - Previous unimodal/multimodal experience
  - Available CPU / GPU resources
- For topics or datasets not in the list:
  - Include a description with links (for other students)

Course Grades



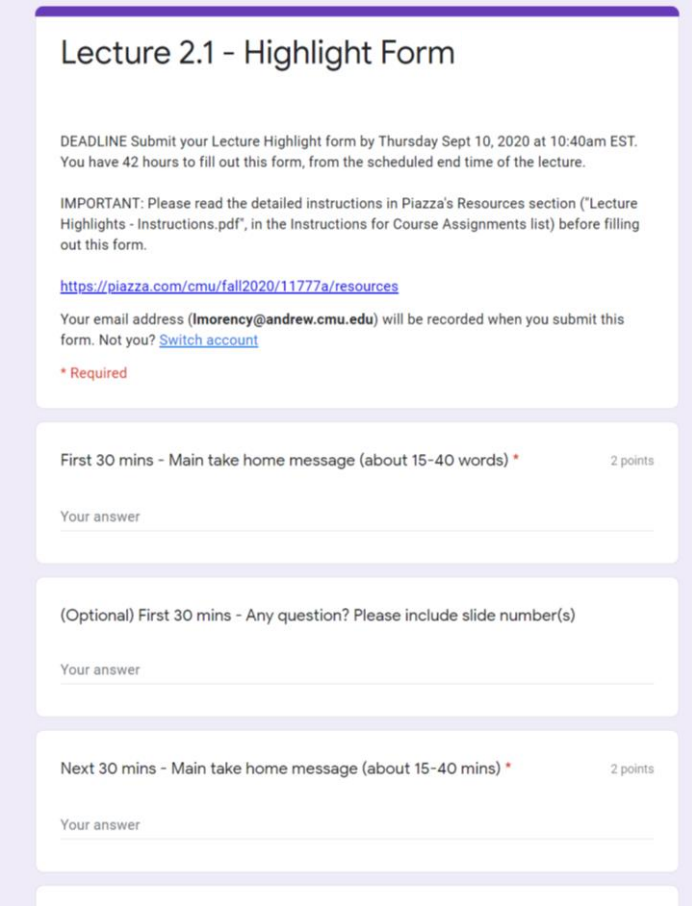
$$
\begin{aligned}
i_t &= \sigma\left(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i\right) \\
f_t &= \sigma\left(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f\right) \\
c_t &= f_t c_{t-1} + i_t \tanh\left(W_{xc}x_t + W_{hc}h_{t-1} + b_c\right) \\
o_t &= \sigma\left(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o\right) \\
h_t &= o_t \tanh(c_t)
\end{aligned}
$$

- Lecture highlights 16%
- Reading assignments 12%

- Project preferences/pre-proposal 2%
- First project assignment 10%
- Second project assignment 10%
- Mid-term project assignment
  - Report and presentation 20%
- Final project assignment
  - Report and presentation 30%

Lecture Highlight Form

Starting Week 2 !!


Lecture 2.1 - Highlight Form

DEADLINE Submit your Lecture Highlight form by Thursday Sept 10, 2020 at 10:40am EST. You have 42 hours to fill out this form, from the scheduled end time of the lecture.

IMPORTANT: Please read the detailed instructions in Piazza's Resources section ("Lecture Highlights - Instructions.pdf", in the Instructions for Course Assignments list) before filling out this form.

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

Your email address (**lmorency@andrew.cmu.edu**) will be recorded when you submit this form. Not you? Switch account

* Required

First 30 mins - Main take home message (about 15-40 words) * 2 points

Your answer

(Optional) First 30 mins - Any question? Please include slide number(s)

Your answer

Next 30 mins - Main take home message (about 15-40 mins) * 2 points

Your answer

Similar to note-taking during lectures

- For each course segment (30mins):
  2 sentences describing the main points

Help you summarizing the lecture

- What is the main take-away message from the lecture
  Short paragraph (15-40 words)

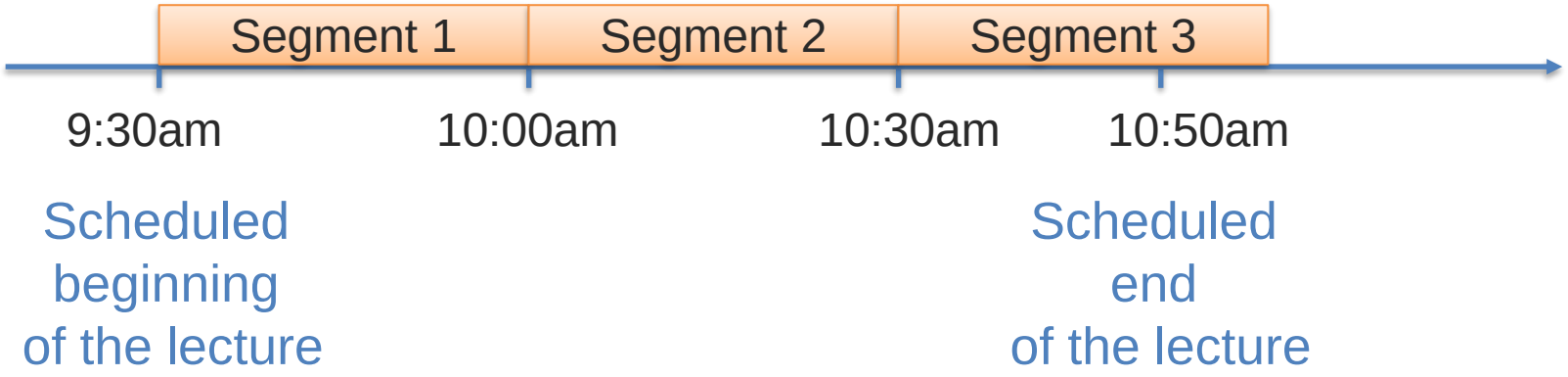
Ask questions about the lecture

- Will be answered either online or at the next lecture

Submitted same day as lecture (before 8pm)

- Students are encouraged to attend lectures in person

Lecture Highlight Form - Segments



- Segment 1 starts at 9:30am, even if the lecture starts slightly later.
- Segment 3 ends whenever the lecture ends
- Slides happening around the segment borders (+/- 5min of 10:00am and 10:30am) can be included in either neighboring segment.

First Reading Assignment – Week 2

- Study groups: 9-10 students per group (randomly, in Piazza)
- 4 paper options are available
  - **Each student should pick one paper option!**
    - Google Sheets were created to help balance the papers between group members
  - Then you will create a short summary to help others [1 point]
- Discussions with your study group
  - Read other's summaries. Ask questions!
  - Write follow-up posts comparing the papers and suggesting ideas [1 point]
    - At least one follow-up post for every paper you did not read

First Reading Assignment – Week 2

Four main steps for the reading assignments

1. **Monday 8pm:** Official start of the assignment
2. **Wednesday 8pm:** Select your paper
3. **Friday 8pm:** Post your summary
4. **Monday 8pm:** Post your follow-up posts

Detailed instructions posted on Piazza

https://piazza.com/cmu/fall2023/11777/resources

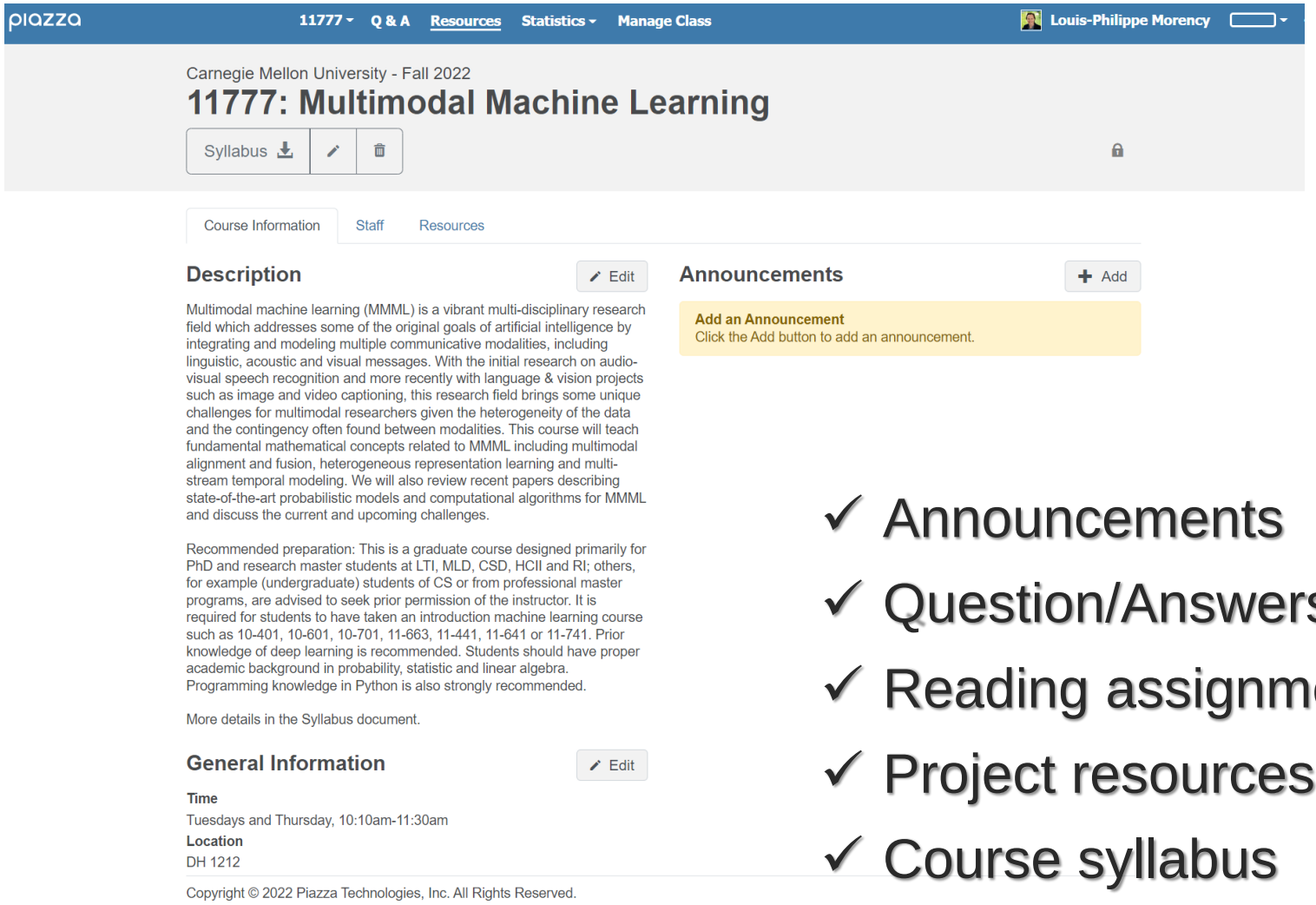
Late Submissions and Wildcards

- Each student has **6** late submission wildcards
  - For lecture highlight forms or reading assignments
- Each project team has **2** late submission wildcards
  - For any of the project assignments
- Total number of wildcards: 8 (6 individual and 2 team-level)
- Each wildcard gives 24-hour extension
  - No partial credits for the wildcards
  - Automatically calculated (no need to contact us apriori)

See details about late submission policy in syllabus

https://piazza.com/cmu/fall2023/11777/resources

Piazza https://piazza.com/cmu/fall2023/11777/info



- ✓ Announcements
- ✓ Question/Answers
- ✓ Reading assignments
- ✓ Project resources
- ✓ Course syllabus