

# How to AI (Almost) Anything

## Lecture 1 - Introduction

**Paul Liang**

Assistant Professor

MIT Media Lab & MIT EECS



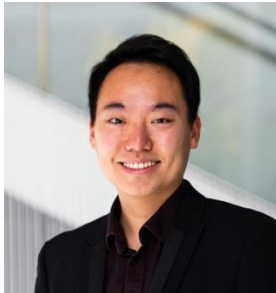
<https://pliang279.github.io>

[ppliang@mit.edu](mailto:ppliang@mit.edu)

 [@pliang279](https://twitter.com/pliang279)



# Your Teaching Team, Spring 2025



**Paul Liang**  
[ppliang@mit.edu](mailto:ppliang@mit.edu)  
Course instructor



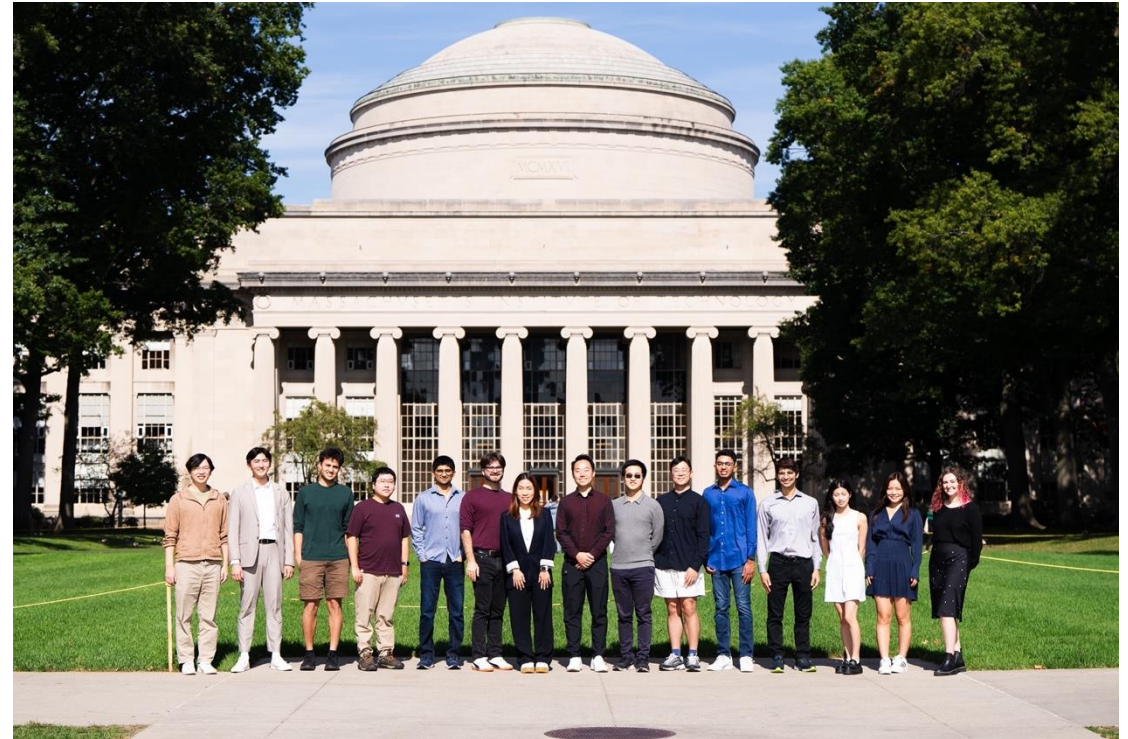
**Chanakya Ekbote**  
[cekbote@mit.edu](mailto:cekbote@mit.edu)  
Teaching Assistant



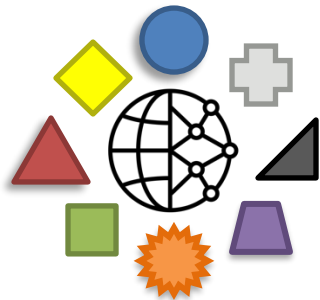
**David Dai**  
[dvdai@mit.edu](mailto:dvdai@mit.edu)  
Teaching Assistant

# **M** multisensory intelligence

Creating human-AI symbiosis across scales and sensory mediums to enhance productivity, creativity, and wellbeing.



## Foundations of multisensory AI



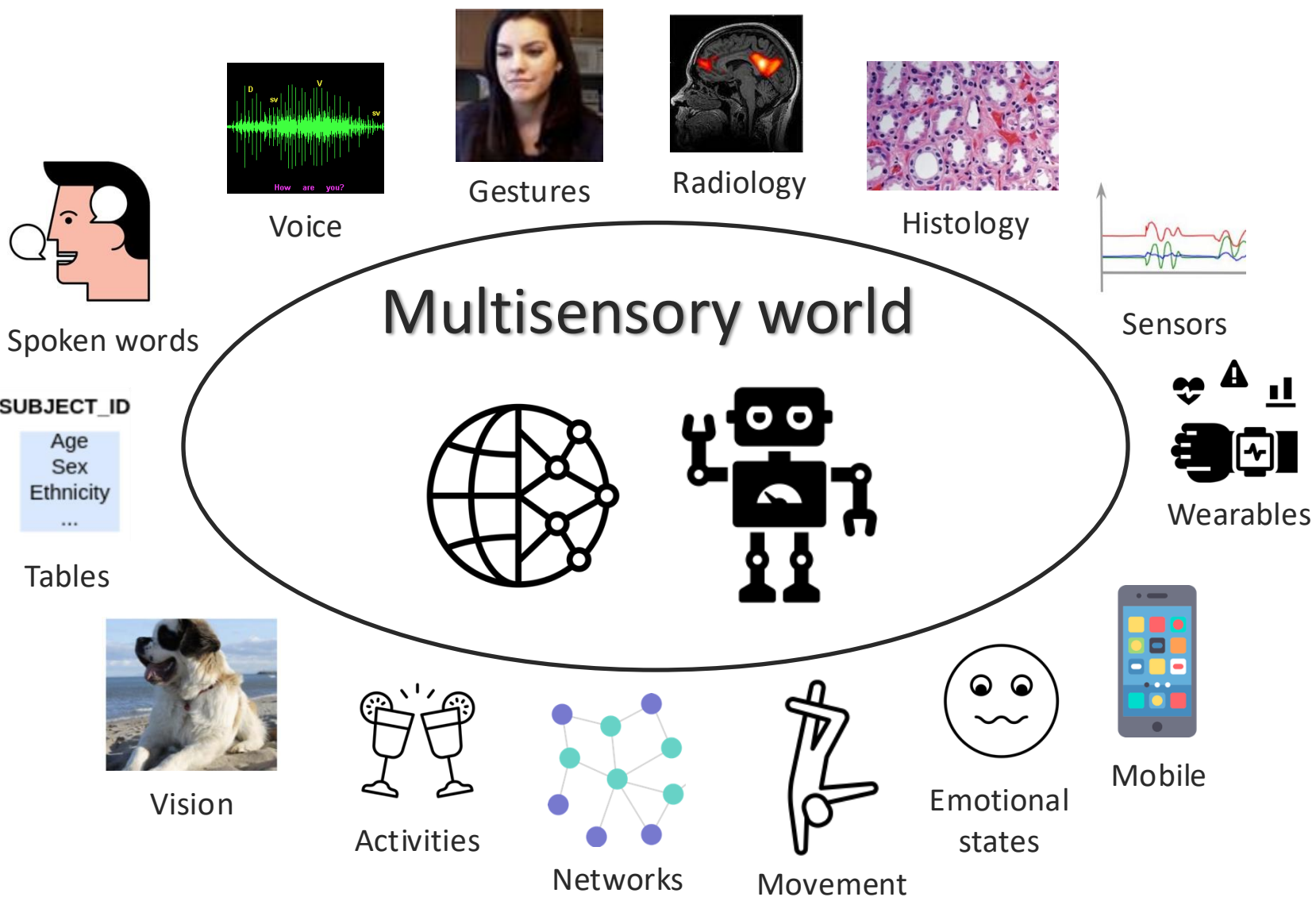
## Enhancing the human experience



## Real-world interaction

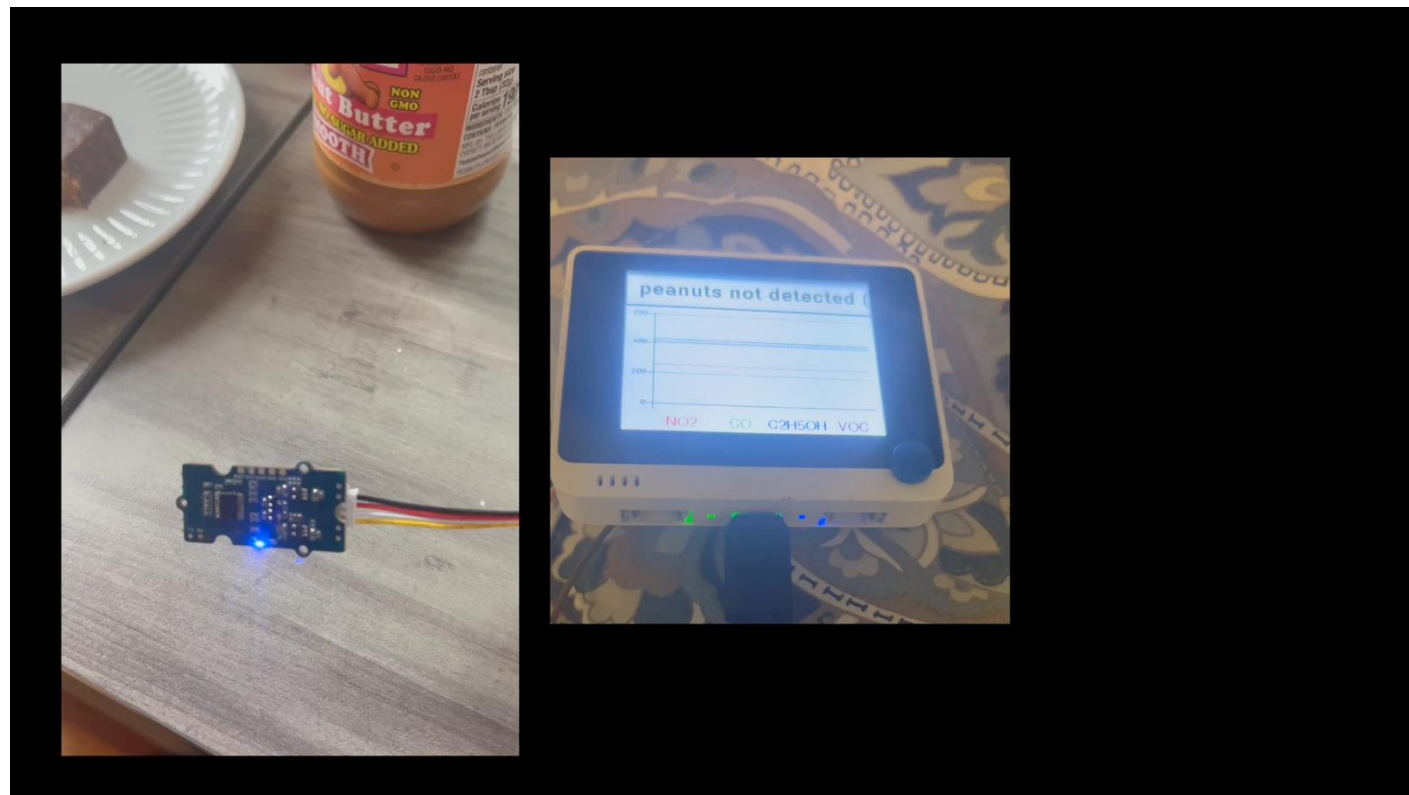
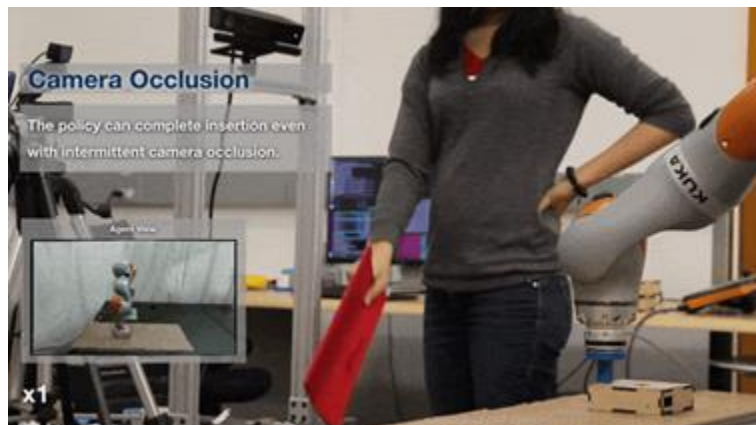
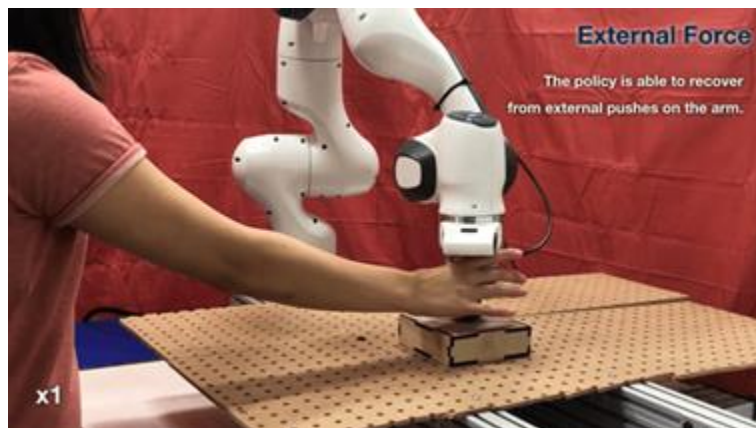


# AI for Anything!



# AI for Physical Sensing

Sensing in physical systems, manufacturing, smart cities, IoT, robotics

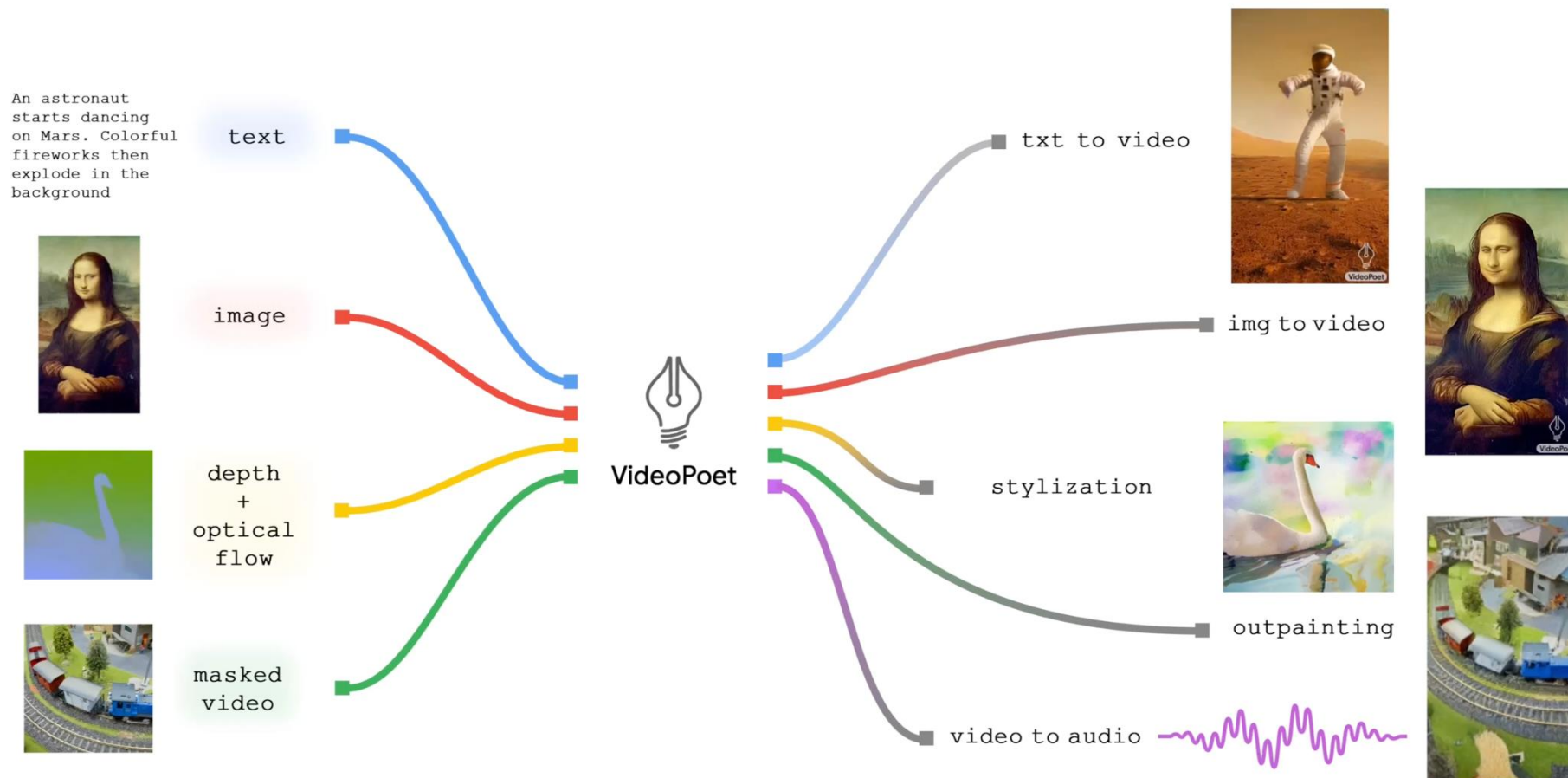


[Lee et al., Making Sense of Vision and Touch: Learning Multimodal Representations for Contact Tasks. ICRA 2019]

[Feng et al., SmellNet: A Large-scale Hierarchical Database for Real-world Smell Recognition. In progress 2024]

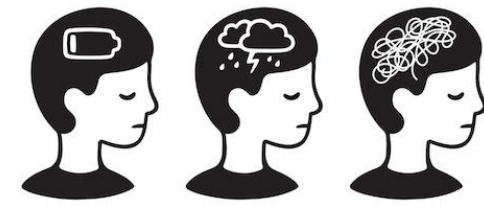
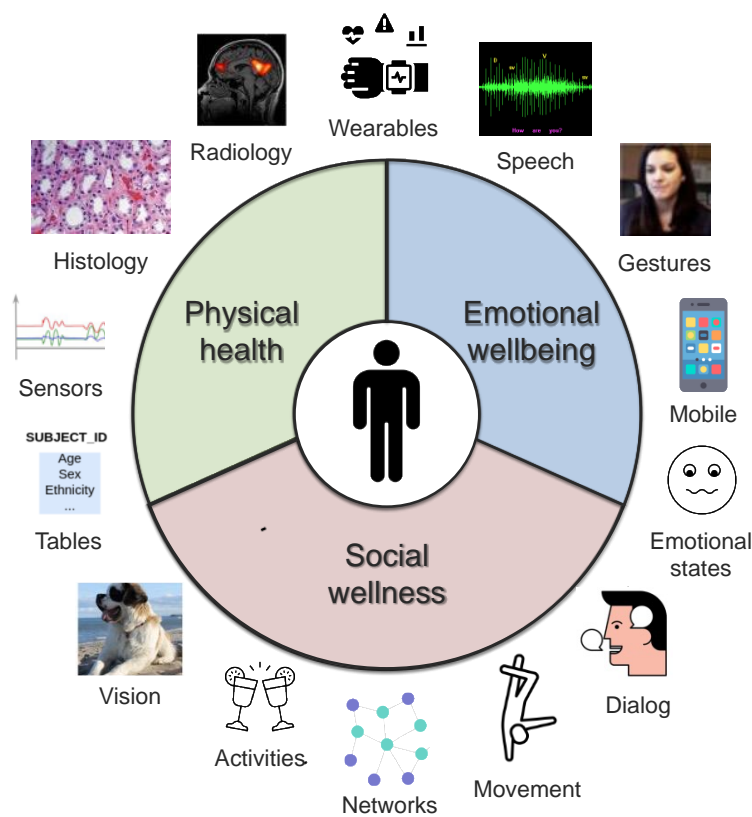
# Multimodal Generative AI

Multimedia, content creation, creativity and the arts



# Holistic Health: Physical, Social, and Emotional

Majority of medical indicators will not be taken in the doctor's office



[Dai et al., Clinical Behavioral Atlas. NEJM AI 2025]

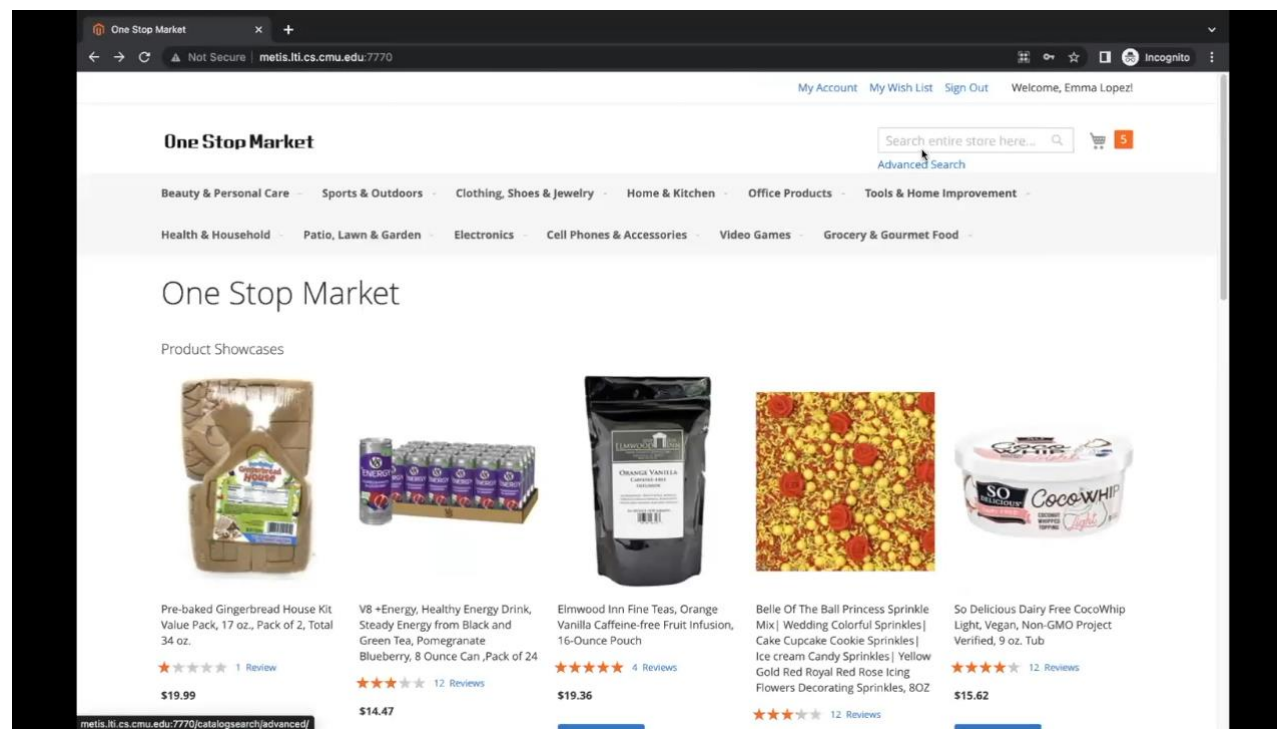
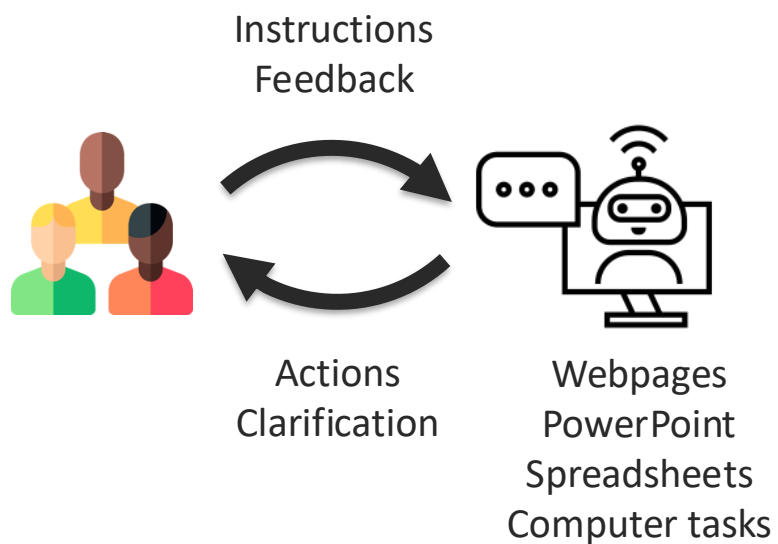
[Hu et al., OpenFace 3.0: An Open-source Foundation Model for Facial Behavior Analysis. In progress 2024]

[Mathur et al., Advancing Social Intelligence In AI: Technical Challenges and Open Questions. EMNLP 2024]

# Interactive Agents

AI agents for the web and digital automation

Example task: Purchase a set of earphones with at least 4.5 stars in rating and ship it to me.

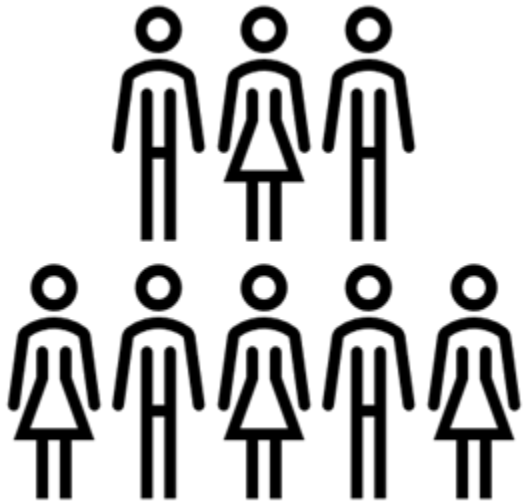


[Zhou et al., WebArena: A Realistic Web Environment for Building Autonomous Agents. ICLR 2024]

[Jang et al., VideoWebArena: Evaluating Multimodal Agents on Video Understanding Web Tasks. ICLR 2025]



# Time for Introductions!



Your name, department and programs

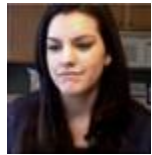
Your favorite modality(ies)!

Previous research experience in AI

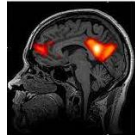
Why are you interested in this course?

# Course Overview

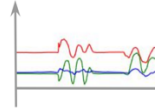
## 1. AI for new modalities: data, modeling, evaluation, deployment



Gestures



Radiology



Sensors



Wearables

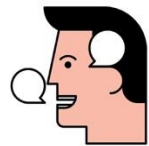


Mobile

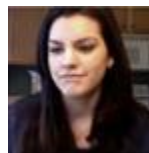


Networks

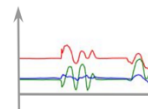
## 2. Multimodal AI: connecting multiple different data sources



Language



Gestures



Sensing



Actuation

# Learning Objectives

- 1 Study recent technical achievements in AI research
- 2 Improve critical and creative thinking skills
- 3 Understand future research challenges in AI
- 4 Explore and implement new research ideas in AI

# Preferred Pre-requisites

- 1 Some knowledge of programming (ideally in Python)
- 2 Some basic understanding of modern AI capabilities & limitations
- 3 Bring external (non-AI) domain knowledge about your problem
- 4 Bonus: worked on AI for some modality

# Course delivery format

- 1-hour lecture every Tuesday
- 1-hour discussion or hands-on tutorial every Thursday
- Reading assignments outside of class
- Significant research project outside of class, with reports and presentations

# Lecture Topics (subject to change, based on student interests and course discussions)

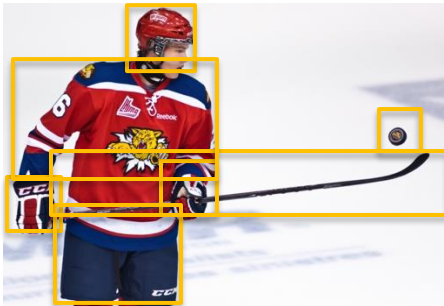
## Module 1: Foundations of AI

Week 1 (2/4): Introduction to AI and AI research

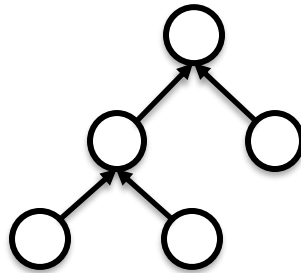
Week 2 (2/11): Data, structure, and information

Week 3 (2/18): Common model architectures

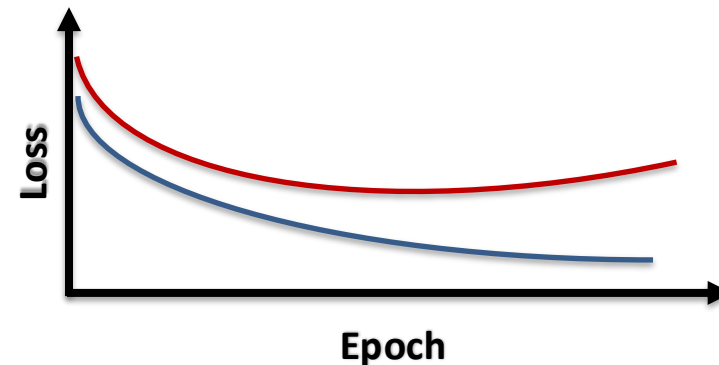
Week 4 (2/25): Learning and generalization



Spatial



Hierarchical



# Lecture Topics (subject to change, based on student interests and course discussions)

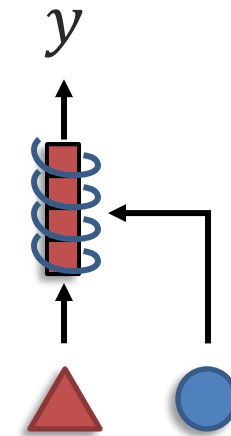
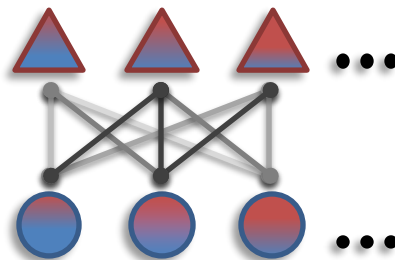
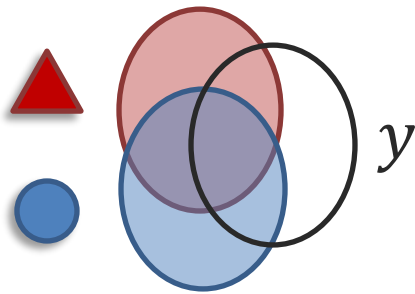
## Module 2: Foundations of multimodal AI

Week 5 (3/4): Multimodal connections and alignment

Week 6 (3/11): Multimodal interactions and fusion

Week 7 (3/18): Cross-modal transfer

Week 8 – No class, spring break



# Lecture Topics (subject to change, based on student interests and course discussions)

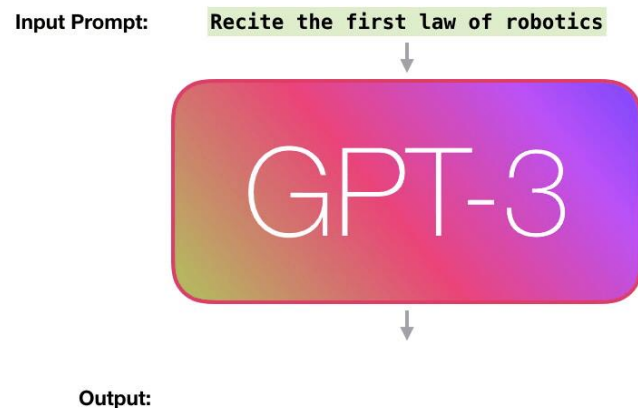
## Module 3: Large models and modern AI

Week 9 (4/1): Pre-training, scaling, fine-tuning LLMs

Week 10 – No class, member's week

Week 11 (4/15): Large multimodal models

Week 12 (4/22): Modern generative AI



*An armchair in  
the shape of an  
avocado*





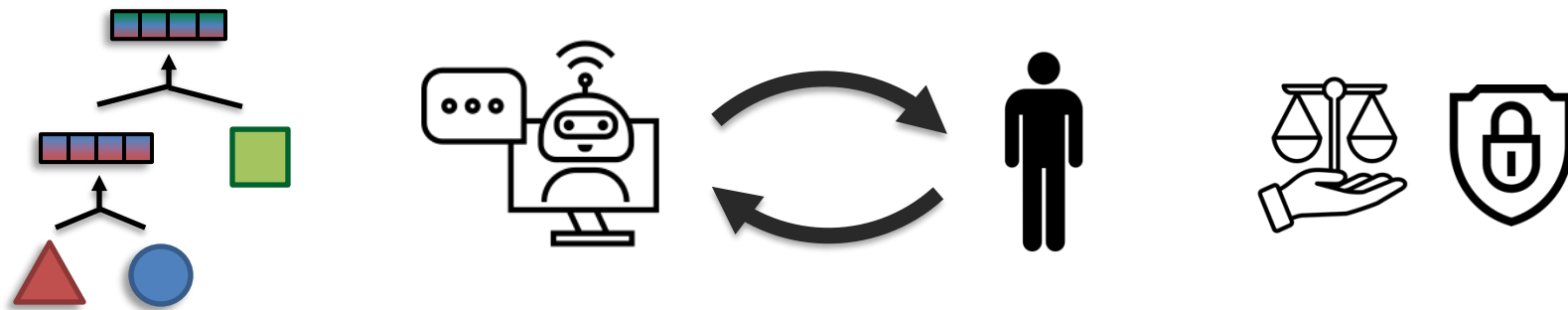
# Lecture Topics (subject to change, based on student interests and course discussions)

## Module 4: Interactive AI

Week 13 (4/29): Multi-step reasoning

Week 14 (5/6): Interactive and embodied AI

Week 15 (5/13): Human-AI interaction and safety



# Grading Overview

- 40% of the grade:
  - Reading assignments
  - Small group discussions
  - Synopsis leads
- 60% of the grade:
  - A high-quality research project:
    - *Proposal with literature review*
    - *Midterm and final reports and presentations*
    - *Bi-weekly updates*

# Reading Assignments

- 7 readings assignments, with usually 2 required papers and some suggested (but optional) papers, and 5-6 discussion probes.
- Three main assignment parts (due Monday night before discussion):
  - **Reading notes:** Read the assigned papers and summarize the main take-away points
    - *Optional: if you have clarification questions about the papers*
  - **Paper scouting:** Scout for extra papers, blog posts or other resources related to these question probes
  - **Discussion points:** Reflect on the question probes related to the reading papers and prepare discussion points.

# Weekly Thursday Class

- Joint portion (about 15 mins)
  - Short presentation presenting the scouted papers and answering student questions about the required papers.
- Separate (TBD based on final enrollment) discussion groups (about 40 mins)
  - 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 (alternate note-taking).

# Discussion Roles

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

1. Short presentation (10-15 mins), done Sunday night - Thursday
  - a) *Answer questions from other students*
  - b) *Summarize and highlight scouted papers*
2. Help with note-taking during discussions

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

1. Note-taking during discussions
2. Summarize discussion synopsis, done Thursday - Monday
  - a) *Merge notes from both groups*
  - b) *Summarize the main discussion points*
  - c) *Organize into an overview schema, table or figure*

# Discussion Topics

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

Week 4 (2/27): Learning and generalization

Week 5 (3/6): Specialized vs general architectures

Week 6 (3/13): Cross-modal transfer

Week 7 (3/20): Large language models

Week 11 (4/17): Large multimodal models

Week 12 (4/24): Modern generative AI

Week 13 (5/1): Human-AI interaction

# Grading Scheme for Readings and Discussions

- Reading assignments 15%
  - 6 points per reading assignment session
    - **1 point** for scouting relevant resources
    - **2 points** for take-away messages from the assigned papers
    - **3 points** for reflections and thoughts on open discussion probes
  - Total 7 reading assignments

# Grading Scheme for Readings and Discussions


- Participation and discussions 15%
  - 4 points per discussion session
    - **2 points** for the insight and quality of the shared discussion points
    - **2 points** for interactivity and participation as follow-up to other's questions and suggestions.
  - Total 7 reading discussions





# Grading Scheme for Readings and Discussions


- Special leads 10%
  - Reading leads:
    - **4 points** for preparing and delivering the presentation at the start of class
    - **1 point** for taking notes during the discussion
  - Synopsis leads:
    - **1 point** for taking notes during the discussion
    - **4 points** for creating the post-discussion synopsis summarizing the take-home messages
  - 1-2 times over the semester for each student


# Other Discussion Roles


 **Scientific Peer Reviewer.** The paper has not been published yet and is currently submitted to a top conference where you've been assigned as a peer reviewer. Complete a full review of the paper answering all prompts of the official review form of the top venue in this research area (e.g., *NeurIPS*). This includes recommending whether to accept or reject the paper.


 **Archaeologist.** This paper was found buried under ground in the desert. You're an archeologist who must determine where this paper sits in the context of previous and subsequent work. Find and report on one *older* paper cited within the current paper that substantially influenced the current paper and one *newer* paper that cites this current paper.

 **Academic Researcher.** You're a researcher who is working on a new project in this area. Propose an imaginary follow-up project *not just* based on the current but only possible due to the existence and success of the current paper.

 **Industry Practitioner.** You work at a company or organization developing an application or product of your choice (that has not already been suggested in a prior session). Bring a convincing pitch for why you should be paid to implement the method in the paper, and discuss at least one positive and negative impact of this application.

 **Hacker.** You're a hacker who needs a demo of this paper ASAP. Implement a small part or simplified version of the paper on a small dataset or toy problem. Prepare to share the core code of the algorithm to the class and demo your implementation. Do not simply download and run an existing implementation – though you are welcome to use (and give credit to) an existing implementation for “backbone” code.

 **Private Investigator.** You are a detective who needs to run a background check on one of the paper's authors. Where have they worked? What did they study? What previous projects might have led to working on this one? What motivated them to work on this project? Feel free to contact the authors, but remember to be courteous, polite, and on-topic.

 **Social Impact Assessor.** Identify how this paper self-assesses its (likely positive) impact on the world. Have any additional positive social impacts left out? What are possible negative social impacts that were overlooked or omitted?

# A Typical Week for Reading Assignments

- Previous Wednesday - @All reading assignment released
- Monday - @All reading assignment due
- Wednesday - @Reading leads make slides for clarifications + scouted papers
- **Thursday** - @Reading leads present slides
- **Thursday** - @All discussion in 2 groups
- **Thursday** - @Synopsis leads take notes with help from @Reading leads
- **Thursday** - @Reading leads submit slides for grading
- Thursday - @Synopsis leads submit 2 sets of notes
- Next Monday - @Synopsis leads merge notes and create coherent synopsis

# Which weeks would you prefer to lead reading & synopsis?

Week 4 (2/27): Learning and generalization

Week 5 (3/6): Specialized vs general architectures

Week 6 (3/13): Cross-modal transfer

Week 7 (3/20): Large language models

Week 11 (4/17): Large multimodal models

Week 12 (4/24): Modern generative AI

Week 13 (5/1): Human-AI interaction

# Research Project

- Similar in spirit to an independent study project
- Project teams of 1 to 3 students
- Final report should be like a research paper
- Expected to explore new research ideas
- Regular meetings with instructors on Thursday

# Research Projects on New Modalities

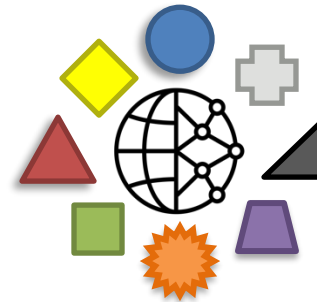
Motivation: Many tasks of real-world impact go beyond image and text.

Challenges:

- AI with non-deep-learning effective modalities (e.g., tabular, time-series)
- Multimodal deep learning + time-series analysis + tabular models
- AI for physiological sensing, IoT sensing in cities, climate and environment sensing
- Smell, taste, art, music, tangible and embodied systems

Potential models and dataset to start with

- Brain EEG Signal: <https://arxiv.org/abs/2306.16934>
- Speech: <https://arxiv.org/pdf/2310.02050.pdf>
- Facial Motion: <https://arxiv.org/abs/2308.10897>
- Tactile: <https://arxiv.org/pdf/2204.00117.pdf>



# Research Projects on AI Reasoning

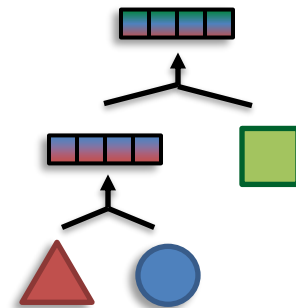
Motivation: Robust, reliable, interpretable reasoning in (multimodal) LLMs.

Challenges:

- Fine-grained and compositional reasoning
- Neuro-symbolic reasoning
- Emergent reasoning in foundation models

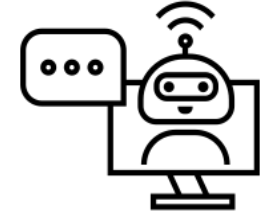
Potential models and dataset to start with

- Can LLMs actually reason and plan?
- Code for VQA: CodeVQA: <https://arxiv.org/pdf/2306.05392.pdf>, VisProg: <https://prior.allenai.org/projects/visprog>, Viper: <https://viper.cs.columbia.edu/>
- Cola: <https://openreview.net/pdf?id=kdHpWogtX6Y>
- NLVR2: <https://arxiv.org/abs/1811.00491>
- Reference games: <https://mcgill-nlp.github.io/imagecode/>, <https://github.com/Alab-NII/onecommon>, <https://dmg-photobook.github.io/>



# Research Projects on Interactive Agents

Motivation: Grounding AI models in the web, computer, or other virtual worlds to help humans with digital tasks.



## Challenges:

- Web visual understanding is quite different from natural image understanding
- Instructions and language grounded in web images, tools, APIs
- Asking for human clarification, human-in-the-loop
- Search over environment and planning

## Potential models and dataset to start with

- WebArena: <https://arxiv.org/pdf/2307.13854.pdf>
- AgentBench: <https://arxiv.org/pdf/2308.03688.pdf>
- ToolFormer: <https://arxiv.org/abs/2302.04761>
- SeeAct: <https://osu-nlp-group.github.io/SeeAct/>

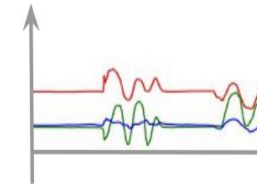


# Research Projects on Embodied and Tangible AI

Motivation: Building tangible and embodied AI systems that help humans in physical tasks.

## Challenges:

- Perception, reasoning, and interaction
- Connecting sensing and actuation
- Efficient models that can run on hardware
- Understanding influence of actions on the world (world model)

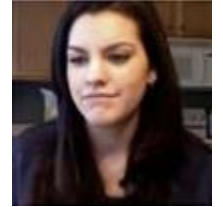
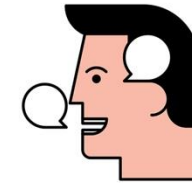


## Potential models and dataset to start with

- Virtual Home: <http://virtual-home.org/paper/virtualhome.pdf>
- Habitat 3.0 <https://ai.meta.com/static-resource/habitat3>
- RoboThor: <https://ai2thor.allenai.org/robothor>
- LangSuite-E: <https://github.com/bigai-nlco/langsuite>
- Language models and world models: <https://arxiv.org/pdf/2305.10626.pdf>

# Research Projects on Socially Intelligent AI

Motivation: Building AI that can understand and interact with humans in social situations.



Challenges:

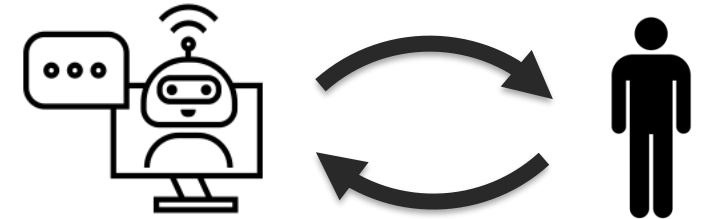
- Social interaction, reasoning, and commonsense.
- Building social relationships over months and years.
- Theory-of-Mind and multi-party social interactions.

Potential models and dataset to start with

- Multimodal WereWolf: <https://persuasion-deductiongame.socialai-data.org/>
- Ego4D: <https://arxiv.org/abs/2110.07058>
- MMTOM-QA: <https://openreview.net/pdf?id=jbLM1yvxaL>
- 11866 Artificial Social Intelligence: <https://cmu-multicomp-lab.github.io/asi-course/spring2023/>

# Research Projects on Human-AI Interaction

Motivation: What is the right medium for human-AI interaction? How can we really trust AI? How do we enable collaboration and synergy?



## Challenges:

- Modeling and conveying model uncertainty – text input uncertainty, visual uncertainty, multimodal uncertainty? cross-modal interaction uncertainty?
- Asking for human clarification, human-in-the-loop, types of human feedback and ways to learn from human feedback through all modalities.
- New mediums to interact with AI. New tasks beyond imitating humans, leading to collaboration.

## Potential models and dataset to start with

- MMHal-Bench: <https://arxiv.org/pdf/2309.14525.pdf> aligning multimodal LLMs
- HACL: <https://arxiv.org/pdf/2312.06968.pdf> hallucination + LLM

# Research Projects on Ethics and Safety

Motivation: Large AI models are can emit unsafe text content, generate or retrieve biased images.



## Challenges:

- Taxonomizing types of biases: text, vision, audio, generation, etc.
- Tracing biases to pretraining data, seeing how bias can be amplified during training, fine-tuning.
- New ways of mitigating biases and aligning to human preferences.

## Potential models and dataset to start with

- Many works on fairness in LLMs -> how to extend to multimodal?
- Mitigating bias in text generation, image-captioning, image generation

# Bi-weekly Project Meetings and Updates

- Required meetings on a bi-weekly basis
  - About 20 minutes per meeting on Thursday afternoon, after class
  - Primary mentor for each team
- Bi-weekly written updates
  - Either Google Slides (preferred) or Google Docs
  - Due Tuesdays at 9pm before the meeting

# Schedule for Bi-Weekly Written Updates and Reports

- Week 3: Proposal presentations
- Week 4: **Proposal report:** baseline results and new ideas
- Week 6: Initial implementation of new ideas
- *Week 8: Spring break (no meetings, no work, relax 😊)*
- Week 9: **Midterm report:** first complete round of results for idea
- Week 9: Midterm presentations
- Week 11: Updated results for research idea
- Week 13: Error analysis, ablations, and visualizations
- Week 14: **Project presentations**
- Week 16: **Final report**

# Course Project Timeline

- **Project preferences** (Due Tuesday 2/11 at 9pm ET) – You should have selected your teammates, have ideas about your dataset and task
- **Proposal report** (Due Tuesday 2/25 at 9pm ET) – Research ideas, review of relevant papers and initial results
- **Midterm report** (Due Tuesday 4/1 at 9pm ET) – Intermediate report documenting the updated results exploring your research ideas.
- **Final report** (Due Tuesday 5/20 at 9pm ET) – Final report describing explored research ideas, with results, analysis and discussion.

# Overall Grades

- First 40% for reading assignments and discussions.
- The second 60% comes from the course project:
  - Proposal report and presentation 10%
  - Midterm report and presentation 15%
  - Final report and presentation 25%
  - Bi-weekly written updates 10%



# 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 (2 per student)
  - 24-hours extension, max 1 per week
- Project report wildcards (2 per team)
  - 24-hours extension, can be used together

# Course Websites

- Course website
  - A public version of the course information
    - *Discussion synopsis will be posted here*
    - <https://mit-mi.github.io/how2ai/spring2025/>
- We will setup canvas for submissions

# Assignments for This Coming Week

No reading assignment this week.

For project:

- Project preference form (Due Tuesday 2/11 at 9pm ET)
  - *To help with team matching*
  - *Google Form link will be available on Piazza*

- Start thinking about what project you want to work on! and potential group mates.

This Thursday: lecture on **how to do AI research**

- From reading papers, to generating ideas, to execution, to paper writing