How to AI (Almost) Anything

Lecture 1.2 – How to AI research

Paul Liang
Assistant Professor
MIT Media Lab & MIT EECS

https://pliang279.github.io
ppliang@mit.edu
@pliang279

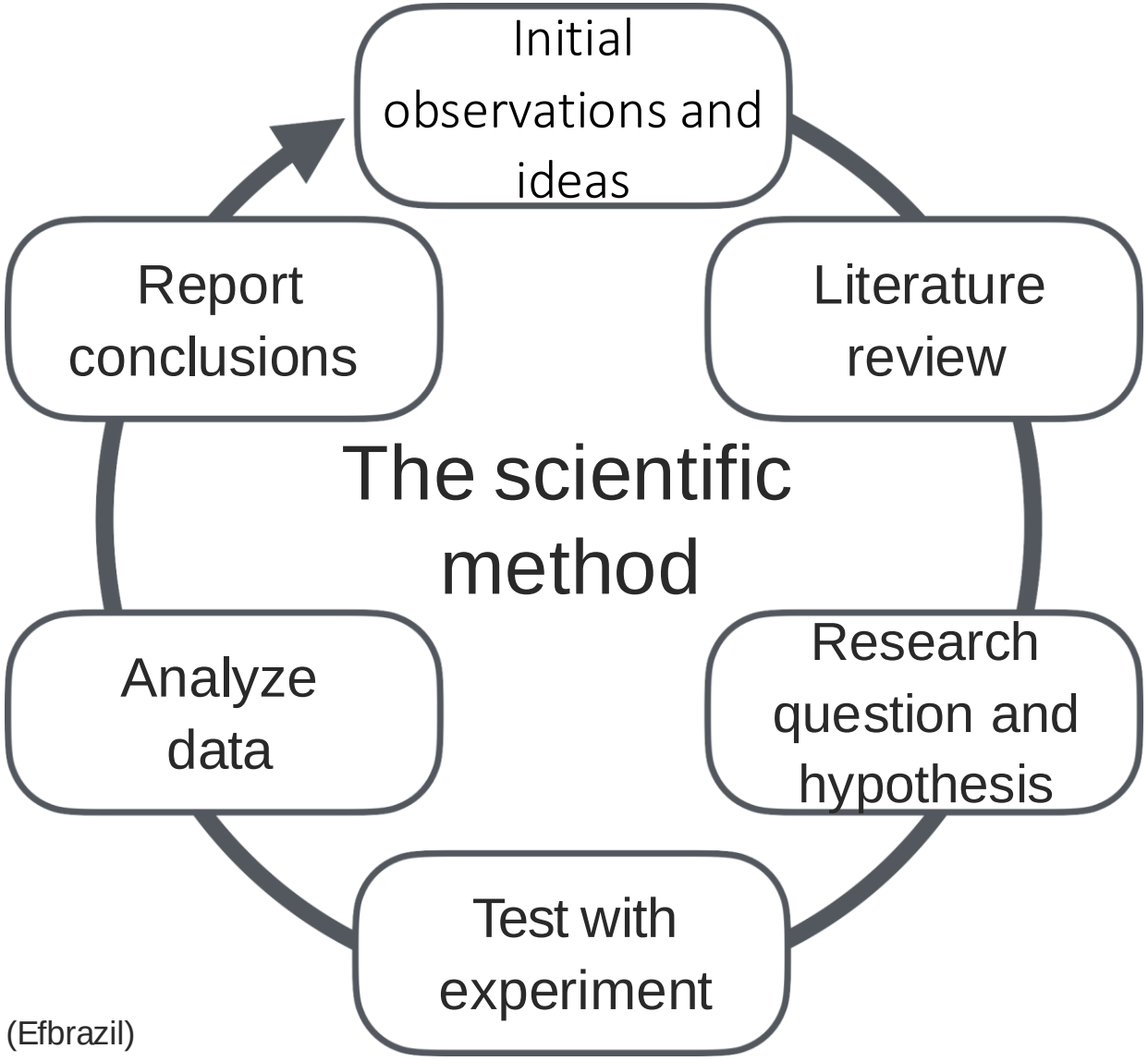
mit media lab

MIT EECS

Today's lecture

1. How to generate research ideas
2. How to read a paper
3. How to execute research ideas
4. How to write a paper

The Research Process



Credit: Adapted From Wikipedia (Efbrazil)

multisensory intelligence

How Do We Get Research Ideas?

Turn a concrete understanding of existing research's failings to a higher-level experimental question.

- **Bottom-up discovery** of research ideas
- Great tool for incremental progress, but may preclude larger leaps

Move from a higher-level question to a lower-level concrete testing of that question.

- **Top-down design** of research ideas
- Favors bigger ideas, but can be disconnected from reality

Bottom-Up Discovery

The proposal report should identify existing methods

The midterm project assignment will enable this bottom-up discovery:
1. Experiment with state-of-the-art models
2. Analyze successes and failures of these models
3. Identify ways you could improve on these failure cases

The final report should propose new ideas that do better

Your research ideas will evolve during the semester!

Top-down Design

Brainstorming: Take the time to brainstorm with your teammates, with TAs and with instructors.

- Office hours with TAs these coming 2 weeks
- Project hours with instructors in the next month

Literature review: The first assignment will allow you to review recent work related to your dataset and your initial research ideas

- When exploring the dataset, you should also expand your research ideas

Midterm and final report focused on incrementally achieving proposed capabilities

Scientific Research Questions and Hypotheses

Research Questions

- One or several explicit questions regarding the thing that you want to know
- Hypotheses are easier to draft with “Yes-no" questions than "how to" questions

Hypothesis:

- What you think the answer to the question may be a-priori
- Should be *falsifiable*: if you get a certain result the hypothesis will be validated, otherwise disproved

multisensory intelligence

Scientific Research Questions and Hypotheses

Are All Languages Equally Hard to Language-Model?

Modern natural language processing practitioners strive to create modeling techniques that work well on all of the world's languages. Indeed, most methods are portable in the following sense: Given appropriately annotated data, they should, in principle, be trainable on any language. However, despite this crude cross-linguistic compatibility, it is unlikely that all languages are equally easy, or that our methods are equally good at all languages.

Cotterell et al. (2018)

What makes a particular podcast broadly engaging?

As a media form, podcasting is new enough that such questions are only beginning to be understood (Jones et al., 2021). Websites exist with advice on podcast production, including language-related tips such as reducing filler words and disfluencies, or incorporating emotion, but there has been little quantitative research into how aspects of language usage contribute to listener engagement.

Reddy et al. (2018)

Exploratory Research Questions

- These questions will be more open-ended
- This is a valid part of research, but you have to be careful about your conclusion claims

For the course research project, exploratory questions are also good options

Beware "Does X Make Y Better?" "Yes"

The above question/hypothesis is natural, but indirect

- If the answer is "no" after your experiments, how do you tell what's going wrong?

Usually you have an intuition about *why* X will make Y better (not just random)

Can you think of other research questions/ hypotheses that confirm/falsify these assumptions

Research Projects on New Modalities

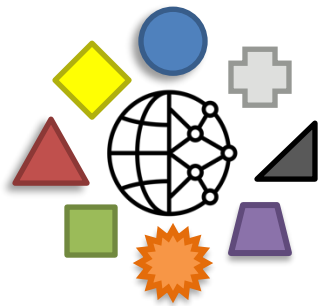
Motivation: Many tasks of real-world impact go beyond image and text.

Challenges:

- AI with non-deep-learning effective modalities (e.g., tabular, time-series)
- Multimodal fusion between deep learning + time-series analysis + tabular models
- Comparing specialized vs general-purpose models
- What type of domain knowledge to include in these models

Potential models and datasets 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



AI for Sensor Data

Motivation: High-frequency, long-range sensing data

Challenges:

- How to tokenize sensor data into semantically meaningful boundaries
- Hybrid signal processing + deep learning methods
- Collecting paired sensor-language data to enable descriptions, captioning, QA, reasoning
- Designing very long-range sensor models with adaptive sensing

Potential models and datasets to start with

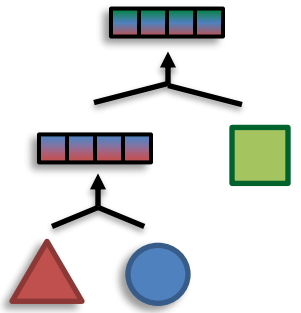
- Search time-series foundation models, changepoint detection.

AI Reasoning

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

Challenges:

- Fine-grained and compositional reasoning
- Neuro-symbolic reasoning, ie. Explicit reasoning
- Emergent (implicit) reasoning in foundation models

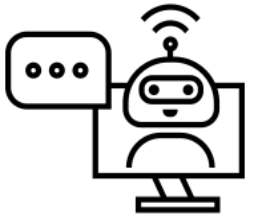


Potential models and datasets to start with

- Many recent reproductions of deepseek/r1/emergent reasoning:
- https://x.com/billyuchenlin/status/1886930173083443233?s=46&t=l93IBUwJM3-kxu9PpUGLpQ
- https://arxiv.org/pdf/2501.19393
- https://arxiv.org/abs/2501.17161
- https://x.com/jiayi_pirate/status/1882839370505621655
- https://x.com/junxian_he/status/1883183099787571519

Interactive Agents

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



Challenges:

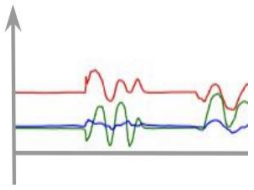
- Web visual understanding is quite different from natural image understanding
- Instructions and language grounded in web images, tools, APIs
- Asking for human clarification, human-in-the-loop
- Search over environment and planning

Potential models and datasets 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/

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)

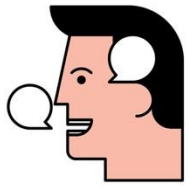
Potential models and dataset to start with

- 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

multisensory intelligence

Socially Intelligent AI

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



Challenges:

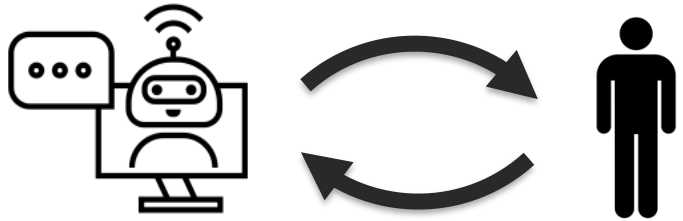
- Social interaction, reasoning, and commonsense.
- Building social relationships over months and years.
- Theory-of-Mind and multi-party social interactions.

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/

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

multisensory intelligence

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

- Better methods for RLHF aligning to human feedback. Different interfaces for human feedback.
- Many works on fairness in LLMs -> how to extend to multimodal?
- Mitigating bias in text generation, image-captioning, image generation

How to do Literature Review and Read a Paper

1. Google scholar
2. Papers with code, Github, Huggingface
3. Recent conference proceedings
4. Blog posts
5. Survey papers, tutorials, courses

Testing Research Ideas

1. Gather and process dataset, visualize data, gather labels, do data splits.
2. Implement the most simple pipeline and get it working.

-> Pipeline = data loading + basic model + eval function + loss/visualization/deployment

3. Change one component of the model at a time, repeat x10 (top-down or bottom-up).
4. Find what works best, and exploit.
5. Scale up experiments, repeat across multiple datasets.
6. Careful ablation studies.
7. Qualitative comparisons and visualizations.
8. Repeat until successful.

How to Write a Paper

1. Prepare a 15min talk (with figures, examples, tables, etc.)
2. Convert the talk into a paper.

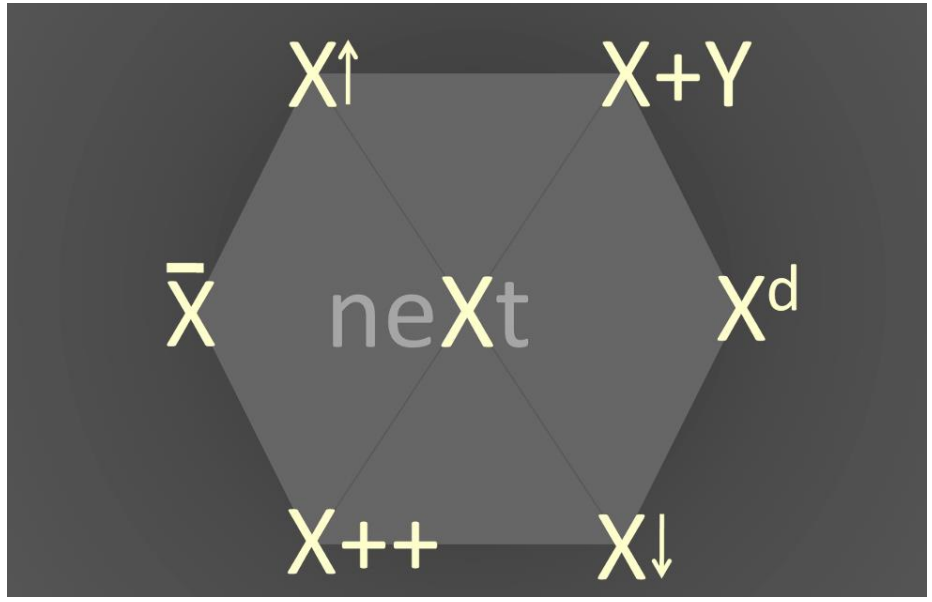
How to Write a Paper

1. Abstract and intro: broad impact, missing gap, intellectual metric (key insight), we propose, summarize results, rehash contributions (1 key contribution, maybe some secondary)
2. Related work organized by themes, end each theme with missing gap
3. Overview figure on page 1/2 for beginner
4. Detailed figure on page 3/4 with components
5. Methods section organized by problem statement, key components, data, training
6. Clear experimental setup with ~3 research questions
7. Big results table on page 5, bold numbers, multiple datasets, error bars
8. Results section organized by answering research questions, and bonus results
9. Some small tables on page 6/7 testing ablations (esp key contribution)
10. Multiple figures on page 7/8 with qualitative comparisons and visualizations

** all figures and tables must have standalone captions
** exactly 8 pages, nothing more and nothing less

More resources

https://github.com/pliang279/awesome-phd-advice
https://github.com/jbhuang0604/awesome-tips
https://www.cs197.seas.harvard.edu/
https://medium.com/spotprobe/the-hexagon-of-ideas-02e5b770d75e



multisensory intelligence

Available Tools

- Use available tools in your research groups
  - Or pair up with someone that has access to them
- Find some GPUs!
- We will be getting AWS credit for some extra computational power
- Google Cloud Platform credit as well

Assignments for This Coming Week

For project:

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

Next Tuesday: lecture on **data, structure, and information**

- Types of data, intuitions on their structure and information, how to model them