Human AI Interaction

CS 485/584

Why did you take the class?

Human-AI interaction – why should you learn about it?

- AI is the new electricity

Human-AI interaction – why bother learning it?

- AI is the new electricity… but how do we use it?

Chinmay's personal experience with AI

Sep 2010: Third week of class.

Theorem. Let $\mathcal{H}$ be given, and let $d = \mathrm{VC}(\mathcal{H})$. Then with probability at least $1-\delta$, we have that for all $h \in \mathcal{H}$,

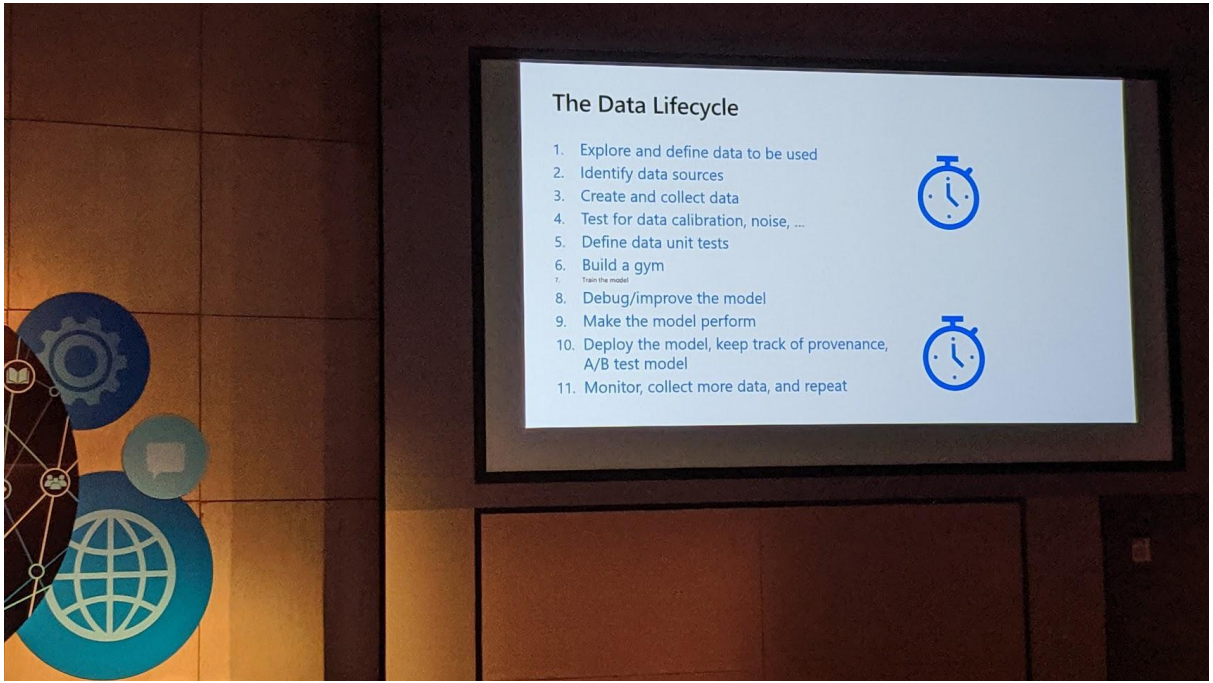
$$|\varepsilon(h) - \hat{\varepsilon}(h)| \leq O\left(\sqrt{\frac{d}{m}\log\frac{m}{d} + \frac{1}{m}\log\frac{1}{\delta}}\right).$$

Thus, with probability at least $1-\delta$, we also have that:

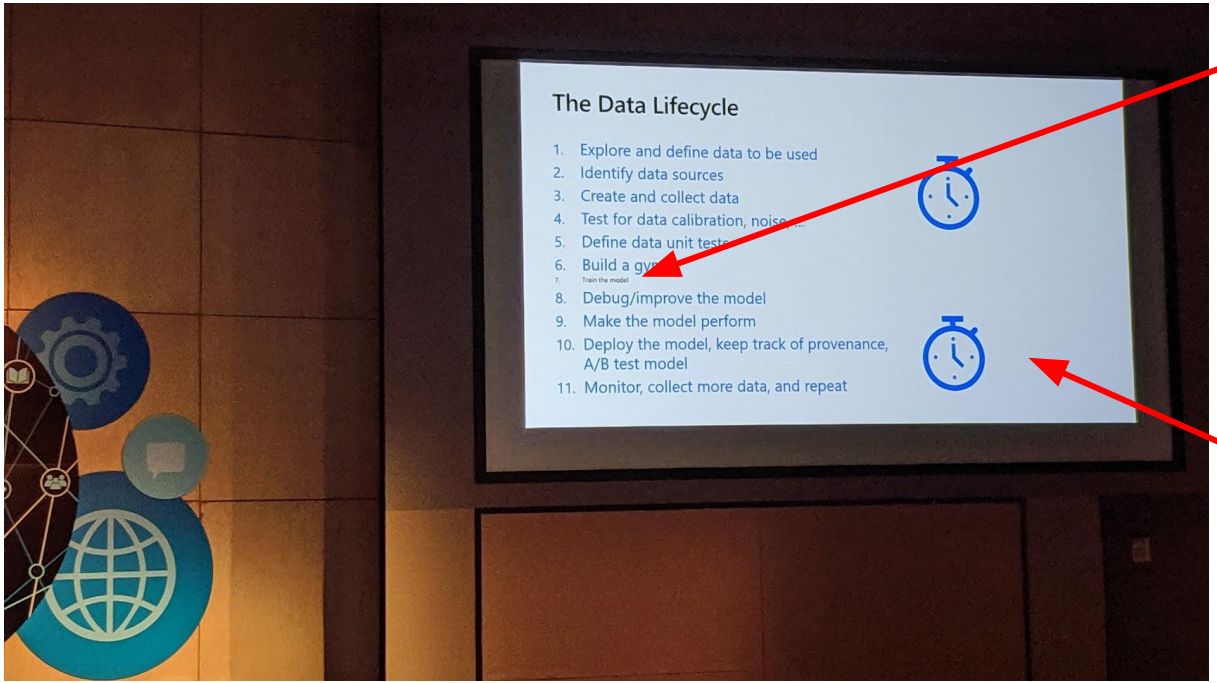
$$\varepsilon(\hat{h}) \leq \varepsilon(h^*) + O\left(\sqrt{\frac{d}{m}\log\frac{m}{d} + \frac{1}{m}\log\frac{1}{\delta}}\right).$$

I've used this beautiful mathematical result exactly once while building interactive AI

In 2019… (by the team lead for Skype)



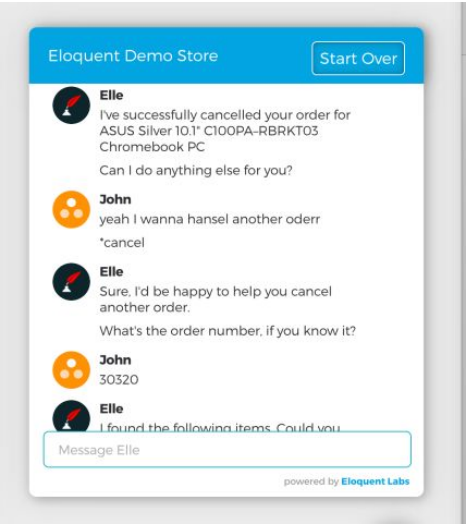
In 2019… (by the team lead for Skype)



Good: train your model is tiny

Bad: where are all the humans???

My experiences have convinced me interaction matters



2016-2019 advising Eloquent Labs (later acquired by Square)

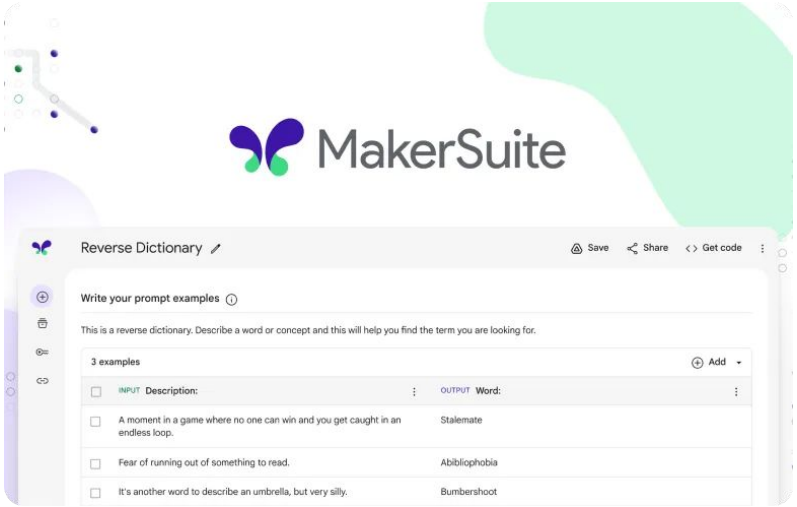
What we thought would be hard:

- Understanding intent
- Mapping intents to actions

What turned out to be harder:

- Did they type accept but mean “except”?
- Should a bot that lets you return items steer you to buy something else?
- Humans change their mind halfway through a conversation… what should we do?

My experiences have convinced me interaction matters



In 2022-2023, as founding member of Makersuite (Google Research)

What we thought would be hard:

- Getting PaLM 2 to follow complicated instructions
- Nudging users to follow ML best practice

What turned out to be just as hard:

- “What do I use this for?”
- “Should my instructions be precise or vague?”
- “How many examples?”

Why we’re doing this course

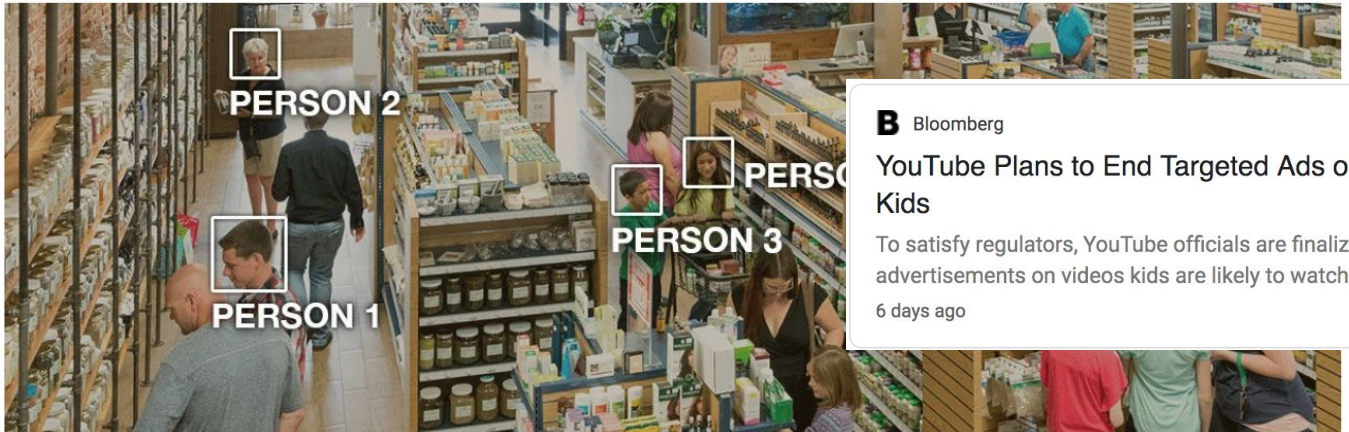
- Most AI/ML courses consider “user-interfaces” or human impact as an afterthought; and focus narrowly on algorithms (and sometimes data)
- My experience working in big companies, small companies, and advising government agencies:
  - Interaction >> modeling
- With you, we are co-inventing a new user-centric way to build AI systems - let’s do it!

Getting it wrong is not an option

Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day

Russian trolls, bots 'spreading discord' over vaccine safety, scientists say

Amazon Pushes Facial Recognition to Police. Critics See Surveillance Risk.



B Bloomberg

YouTube Plans to End Targeted Ads on Videos Aimed at Kids

To satisfy regulators, YouTube officials are finalizing plans to end "targeted" advertisements on videos kids are likely to watch, according to ...

6 days ago

What will you get out of this course?

- **Design**: Systematic techniques to design new [interaction + AI] systems (e.g. a notetaking copilot, a new game, an app to reduce food waste, or whatever you want!)
- **Implementation**: Dive deep into new large foundation models like LLMs, text-to-image etc. What makes these models different, and how do they change AI practice?
- **Evaluation** of interactive AI: How do you evaluate if your AI is useful, harmless, and fair?

What should you expect to do in this course?

- **Thinking it through**: Readings + lectures, quizzes
- **Doing it through**: Assignments and projects
- **Teaching it through**: In class discussions, debate, etc.

Course values

- **Honesty:** your work is yours, acknowledge help, don’t cheat
- **Try your best:** but ask for help if you are stuck
- **Collaboration:** we learn best when we work together

How this course is evaluated

- Assignment 0: 1% [Congratulations!]
- Quizzes: 14% (7 quizzes total; 2% per)
- Individual projects: 30% (2 projects; 15% per)
- Group projects: 30% (2 projects; 15% per)
- “Attendance”: 5% (only *active* participation counts)
- Final project: 20% (group work)

\+ Lots of opportunities for extra credit

No final exam. No midterms.

You do have a final project!

Do you know enough programming?

You need to know some Javascript and Python

Example tasks:

Javascript: Click a button, show a modal

Python: Take data as a csv, and transpose all rows into columns

Let’s begin

We will focus on machine learning today, other AI in the future

A learning machine: A “machine” that is able to improve based on past experience without explicit human programming on how to improve each time.

Machine learning = techniques to help computers complete tasks without explicitly telling them how.

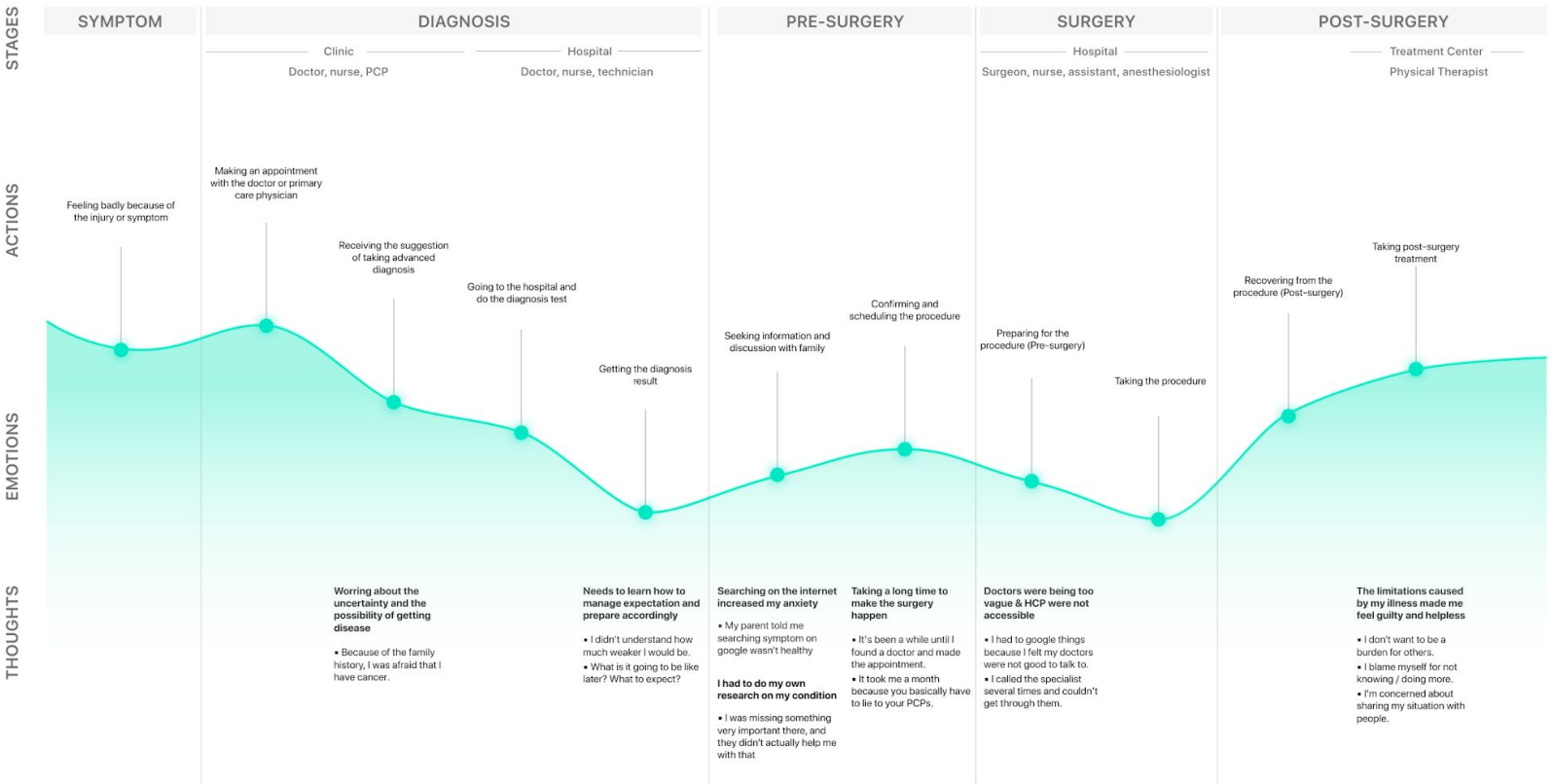
First STEP: Question assumptions

- Most AI starts with lots of hidden assumptions. “Automate customer conversations”.
- Starting today, question:
  - **S**takeholders: who is this for? Who else is affected?
  - **T**ask: is this the right task to automate/augment?
  - **E**xperience: how will we learn how to automate/augment this task?
  - **P**erformance: how should we measure if we are good?

Next STEP: Journey mapping

PATIENT JOURNEY MAP

The end-to-end emotional, behavioral, mental actions in the surgical process



Journey map: What we'll do now (Part 1)

1. (7 minutes) Make your own personal journey map -- alone
2. (3 minutes) Label on your map:
    a. Where did AI come in?
    b. Where should it have?
    c. Where should it not have
3. (5 minutes) Discuss with 3-4 neighbors
    a. What do you see as common themes?
    b. How did your labels differ from others?
4. (5-8 minutes) Share with the class:
    a. common themes
    b. Major points of difference

Analysis (Part 2)

1. (4 minutes) Discuss: The common themes where AI does or should come in:
    a. how might it fail?
    b. What concerns do you have?
2. (4 minutes) Alone: How might you fix one of these concerns
    a. Can you create a purely technical solution?
    b. Can you create a solution part-technical, part people?
3. (5 minutes) Discuss: Solutions
    a. What do you see as common themes among solutions?
4. (5-8 minutes) Share with the class:
    a. common themes
    b. Unsolved challenges