

CS698Y: Human-AI Interaction

Lecture-1

Welcome to the first edition!

What : CS698Y: Human AI Interaction

When : WF 1030-1200

Where : KD103

Who : Dr. Sruti S Ragavan

Show of hands

- Undergrad / graduate students?
- HCI/Design/CogSci backgrounds?
- AI / ML backgrounds?
- Comfortable coding?

More show of hands

- Have social media accounts?
- Used ChatGPT, Gemini or other GenAI tools?
- Use Google Maps?
- Use flight price insights (increase/decrease/right)
- Use train ticket confirmation predictors?

What do you like most/least about ChatGPT?

What do you like most/least about news
Google suggests?

Consider this...

- Most of them are pretty accurate models
 - Even ChatGPT says reasonable things most of the time
 - Google news is great in its predictions
 - Sometimes, a little too much to the point of being creepy
- Why do we still have problems?

The need for human-centric AI

- Model-centric AI
 - Improve accuracy / recall / F1 score
 - Above getting models to be accurate & spit out results
- Human-centric AI
 - Looks at what a human needs
 - Models are just one part of a workflow
 - Other issues → user needs, goals, skills, privacy, ...

What is Human-AI Interaction?

- Subfield of Human-Computer Interaction focused on interactions between humans and AI
 - Interactions = learning, using, building, long-term effect, ...
- Concerned with how AI changes interfaces, tasks, and user expectations and behaviors
- Design and evaluate systems that integrate AI meaningfully into human workflows

Why this course?

- AI/ML is everywhere—but people are the ones using it
- Most failures of AI in the wild are not algorithmic
- Designing AI systems = designing with and for humans
- This course focuses on making AI usable, useful, and responsible

Why AI specifically?

1. Paradigm shift:
 - Users tell computers → computers predict/suggest/act on our behalf
2. ML ≠ deterministic software
3. Long-term impact
 - Having to decide → execute decision suggestions
 - Having to learn/compose → answers directly provided
 - ...
4. Raises new questions:
 - How to design, evaluate, and understand interfaces
 - Who is "in control"? What does "good interaction" mean?
 - We must design for uncertainty, error, trust, ethics, skills of people

Warning: HAI \neq Designing UI wrappers for AI

- Just like HCI does not equal designing interfaces
 - Or physics is about solving equations
 - And computing is about just coding
- HAI is about:
 - Who does what? (Division of labor, partnerships, collaborations)
 - When? (Who initiates, at what times/moments)
 - How? (Modes of interaction, errors/uncertainty, feedback)
 - Why? (Purpose, values, ethics)
- Users come with goals, contexts, skills, and constraints
- AI is a *tool*, not an end

Key Questions in Human-AI Interaction

- Trust: When and how do users trust AI? Too much / too little?
- Control: Who initiates actions – human or AI?
- Transparency: Does the system explain itself clearly?
 - Important to understand how to work with the system
 - Important to debug
- Accountability: Who is responsible for AI outcomes?
- Usability: Is the system helpful and usable for non-experts?
 - Important for maximizing impact
 - Important also for not leaving people out!

Designing for these is hard

- How much autonomy to give the AI?
- How to provide users the agency to decide when AI is right/wrong?
- How to communicate uncertainty?
- How to support correction and feedback?
- How to design for edge cases and failure modes?

Thought experiment

- Suppose Pingala suggested course recommendations to you...
- Do you think it would be useful?
- Assume high accuracy on hindcast data, would you simply accept the recommendations?
 - If yes, why?
 - If not, why not?
- What would make you accept the suggestions?

Real world examples: Healthcare diagnostics

- Doctors use AI for help with diagnostics
- If AI analysis of a patient's reports says "not a cancer" (confidence=0.8712)
 - Should the doctor simply accept it or recommend further tests?
 - What about insurance costs of tests?
 - What about costs of treatments in case of false negative?
 - Who should deal with consequences of misdiagnosis?
 - In some cases, doctors can be sued!

Real world examples: Loan underwriting

- People get loans (housing, education, credit cards...)
- Algorithms decide whether to approve / reject a loan application
 - Approved, but ends up defaulting?
 - Rejected, but could affect overall credit scores?
 - Can even be sued, in some countries? (In India, RTI!)

Real world examples: Self-driving cars

- Who's responsible for accidents?
- What about drivers falling asleep?
- What about driver reflexes failing over time?
- What about drivers not paying attention?
- Job losses?

Real world examples: Robot servers

- Human connections?
 - Smile, greetings, service, tipping
- Spooky?
- Can it interrupt conversations/overhear?
- Loss of jobs?



As engineer, it is important to...

- Understand these details about the real-world where AI systems are deployed
- Design interfaces that communicate appropriately
 - Instead of “Not cancer” (confidence=0.8342)
 - How about: “Very likely not cancer”,
 - Run a biopsy, but not super urgent.
 - Goes from predictor to a priority optimizer.
 - Changes the entire way AI systems are built, framed, marketed, used, thought of as

HAI has been growing in the last 10 years

- Human-centricity: how can AI best help humans?
 - Keeping their needs, contexts, wants, skills in mind
- Large policy, tech organizations in the race
 - Google PAIR
 - Microsoft HAX toolkits
 - IBM Human-AI context toolkit
 - ...

What to expect?

- How to understand people's workflows
- Identify where AI would fit (and where not)
- How to design, evaluate, and integrate AI into real-world workflows
- How to account for trust, explanation, feedback, uncertainty
- Mix of HCI, Software engineering, ML techniques

What this course is not

- Not a theory-of-AI or algorithmic ML course
- You will not learn anything about patterns, algorithms, gradient descent, etc. No calculus/math either.
- Not a UI design course either
- It's about how ML meets human work and reasoning

Background requirements

- ML basics – yes
- Usability design – not assumed, but some idea is good.
- Programming – yes
- We will build gradually, covering broad pre-req across all three areas.

Logistically..



- 9-credit grad level courses
- Expect an average of 8-9 hours a week workload
- Lectures, occasional readings, assignments, quizzes
- All homeworks / slides posted to HelloITK same day as lecture
- A few lectures will be delivered by Dr. RS Sharma
 - Technocrat, thoughtleader in digital public goods
 - Former CEO UIDAI, NHDC, CoWIN for vaccine tracking, ABHA, ...
 - Instrumental in DPDP act, among other things
 - 27th August: Privacy by design in large digital systems

Grading policy

- In-class assignments periodically (10%)
 - Unannounced; Acady quizzes, polls, worksheets, etc.
- Quizzes (15%)
 - Usually announced, based on lecture and readings
- Home works (20%)
 - In pairs, involves a combination of programming, model building, design
- Mid-sem exam (25%)
- Final project, in teams of 2-3 (30%)

Attendance

- Acadly attendance mandatory
- Does not count for grades
- But continued absence has to be reported to DoSA
- Set up the app, **MARK YOUR ATTENDANCE NOW**
- If you are not setup on HelloIITK / Acadly, please talk to me after this class

Next week...

- Principles of Human-AI interaction
- Basics of AI

Questions?