CS 395T: Communicative and Grounded AI Agents

Fall 2025

Today: Introductions

- Who am I?
- Intro and Overview
  - Goals
  - Topics
  - Presentation
  - Project
  - Participation
- How to read and present papers
- How to find papers
- Logistics

Who am I?

- Assistant Professor in Computer Science
- Developing AI agents that intelligently communicate and collaborate with people and each other
  - Multi-agent communication/collaboration (debate, distillation, pragmatics, persuasion)
  - Grounding (vision-language, text-to-action, text-to-code)
  - Handling uncertainty, ambiguity, and underspecification
- CS and CogSci
  - Especially linguistics

Overview

- Course goals:
  - **Understanding** the fundamental questions and challenges in developing agents.
  - An ability to **find, assess, and read cutting-edge research**.
  - An ability to **situate research in historical context**.
  - Brainstorming, developing, and executing an exciting course project in line with course topics.

Overview

- Evaluation
  - Presentation
  - Project
  - Participation

Topics covered/Schedule

Weeks	Topic
1	Intro + LLM background
2-3	LLMs for Action (semantic parsing, tool use, program induction/skill learning, optimization, uncertainty and calibration)
4-5	Multi-agent (orchestration, debate, pragmatics)
5	Embodied/multimodal (VLMs and VLAs)
6-7	Project pitch
7-8	Embodied/multimodal (visual programs code as policy, robotics)

Topics covered/Schedule

Weeks	Topic
9	Computer use (web agents, code agents)
10	Project check-in
10-11	Information agents (deep research, ambiguity, information-seeking)
12	Flex/special topics
13	Project check-in
13-14	Safety/Ethics
15-16	Wrap up/summary, final presentations

About the schedule

- The schedule might change slightly
- Papers might be adjusted
  - I will let you know 2 weeks before you present whether that's the case
- You should feel free to suggest changes
  - Please let me know 1 week in advance via email
  - I will check and confirm

Presentation

- You will each be asked to choose and present on a topic
- How this will work:
  - I have released a topic list and schedule
  - Rank your top 5 days and send to me via email with subject line “CS395T Topic: <Name>”
  - In this email, add 1 sentence about your interests in the course.
  - I will try to match everyone to their highest-ranked available topic. If there is no such topic available, I will allocate topics based on interests.

Presentation

- Presentations will be a reading group format
  - Presenter's job is to be the expert in the room
  - Assume everyone else has read the paper (this is required) but not necessarily that they have understood everything
- Presentations should be uploaded after the class
  - Powerpoint or Google Slides is fine
- Presentations: content matters more than aesthetics
  - That said: they should be digestible

Picking a presentation day

- Each day is assigned multiple papers
  - Some more than others (Note: pick your day by interest
  - Depending on enrollment/interest/number of papers, I may split some of the days between two people
- There are more students than slots, SO
  - I will try to fairly allocate students to slots based on how many papers there are and how content-heavy the papers are
  - Everyone will present at least once
  - Each day will have a secondary presenter
  - Secondary is backup (e.g. if the primary presenter is unexpectedly ill, etc.)
  - If you are primary presenter, please share your slides with me in advance of class, that way I can share them with the secondary presenter if needed
  - If you are secondary presenter, you should plan to read and understand all the papers
  - Secondary presenter should plan to be active during the discussion
  - If you are splitting your presentation day, I will likely assign you as secondary presenter twice

Project

- The main object of this course
- Should be driven by what you want to work on
  - That said: I am happy to suggest ideas
- Can be done together or in groups
  - No more than 3 people per group
  - Each member should contribute equally; group projects will need to include a breakdown of contributions in the final submission
- Finding a group
  - If you already know people, great! Let me know your group and project area asap.
  - If not, I will send around a topic interest form and try to match you
- If you don't want to be in a group
  - Let's chat. Groups are preferred but not strictly required.
- Ambitious project goal: develop the seed of something that can be a conference paper
  - Note: this is not the expectation

Choosing a Project

- Start thinking about this now
- Resources
  - Look at the course schedule: are there any topics that interest you
  - Bring in your outside interests!
- Project can be open-ended; if you are not sure about your topic, ask me if it's a good fit!

Project Timeline

- Start picking a topic
- End of September: project pitches
  - 5-8 slides pitching your project idea
  - Peer review: others will give you feedback on your project (redteaming)
  - Preparing response to redteaming and timeline
- End of October
  - Mid-term check in: give updates on your progress
- End of November (before break)
  - Pre-break check in
- Last week of class
  - Final presentations
- During finals
  - Final writeup
- Goal with all these check-ins: keep your projects on track and give you chances to get feedback

Participation

- One of the course goals: foster a strong discussion
- Being an active audience
  - Ask questions!
- Do's and Don't
  - Quiz questions vs. understanding questions
  - Balancing critique with understanding
  - Major Do: Making connections
- Laptops
  - No official policy on this
  - Guarantee: you will have a more engaged experience if you don't use one

How to read a paper

- Seems simple (hopefully you all know how to read), but…
- Some guidance
  - Papers are not like books
  - Do multiple passes
    - Skim first: try to first identify main points
    - Then come back to fill in the details
  - For reading group purposes
    - Start outlining your slides in the first skim
    - Would recommend skimming all the papers in your batch before starting
      - Think about: How do these fit together? What would you like to highlight?
    - Then: Come back and fill in the details (don't forget to do this part)
- Read curiously!
  - If you have questions, find answers (probably, others will have similar questions)

How to present a paper

- Background
  - Give the context for the paper (a few slides at least)
  - Some days will have entire papers as background
  - Think about whether any key background/context is missing and add that
- Technical papers: Clearly describe:
  - What problem are they solving?
  - Why should we care about it?
  - How did they solve it (method section)?
  - How do they evaluate their method (baselines and metrics)?
  - What were the results?
  - Why does their method work (analysis)?
- Pay attention to this structure! (It will come in handy when you write your reports)

How to find papers

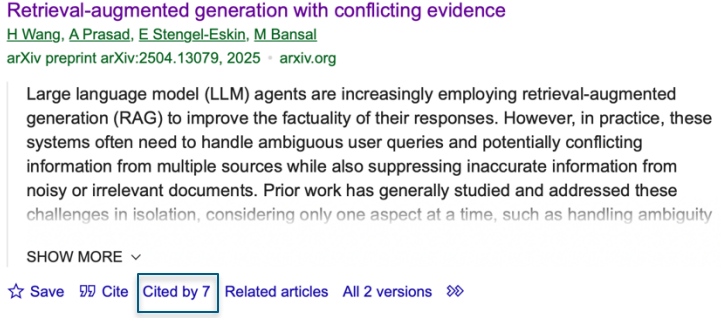
- When reading curiously, you should be able to find related papers
- Backward references and forward references
- Google scholar is your friend
- Generally speaking: LLMs are not your friend here (low recall, hallucinations)
  - Would not recommend unless you know what you are looking for (i.e. you can verify)

References

Anthropic. Claude can now search the web, 2025. URL https://www.anthropic.com/news/web-search.

Akari Asai, Zeqiu Wu, Yizhong Wang, Avirup Sil, and Hannaneh Hajishirzi. Self-RAG: Learning to retrieve, generate, and critique through self-reflection. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=hSyW5go0v8.

Isabelle Augenstein, Timothy Baldwin, Meeyoung Cha, Tanmoy Chakraborty, Giovanni Luca Ciampaglia, David Corney, Renee DiResta, Emilio Ferrara, Scott Hale, Alon Halevy, Eduard Hovy, Heng Ji, Filippo Menczer, Ruben Miguez, Preslav Nakov, Dietram Scheufele, Shivam Sharma, and Giovanni Zagni. Factuality challenges in the era of large language models, 2023. URL https://arxiv.org/abs/2310.05189.



Retrieval-augmented generation with conflicting evidence

H Wang, A Prasad, E Stengel-Eskin, M Bansal

arXiv preprint arXiv:2504.13079, 2025 · arxiv.org

Large language model (LLM) agents are increasingly employing retrieval-augmented generation (RAG) to improve the factuality of their responses. However, in practice, these systems often need to handle ambiguous user queries and potentially conflicting information from multiple sources while also suppressing inaccurate information from noisy or irrelevant documents. Prior work has generally studied and addressed these challenges in isolation, considering only one aspect at a time, such as handling ambiguity

SHOW MORE

☆ Save Cite Cited by 7 Related articles All 2 versions »

Presentation tips

- Limit the amount of text on a slide
  - Generally, slides are visual tools (this is a bad slide)
- A good rule of thumb: 1 minute per slide
  - Plan in pauses for questions
  - Some slides may be more/less (e.g. a slide with many results might need more time)
  - Practice your presentation to get timing right
- Structure your slides (no paragraphs)
- Incorporate and explain figures (draw simple figures if needed)
- Incremental reveals!

Presentation tips

- Some days have many papers
  - Decide what kind of paper each paper is
- For example:
  - If it's an older more foundational paper, focus more on the motivation/influence on other papers as opposed to the results (which might be less impressive now)
  - Read your papers together to try to pull out important common themes, then go back and highlight those ideas
  - For a lot of days, I have chosen papers with this in mind; some of the older papers introduce ideas/tasks/themes that are important for newer papers

Discussion time!

- Let’s hear from you
- No wrong answers

Who are you?

- Who here is a:
  - CS graduate student
  - Linguistics graduate student
  - Neither

Who are you?

- Who here has:
  - Taken an NLP course before
  - Taken an AI or RL course before
  - Done a project with LLMs
  - Used LLMs in their daily lives
  - Used LLM agents in their daily lives

What is an agent?

- What are key features of an agent?
- How is an agent different from a model?

Communication

- What are some examples of communication?
- What are the advantages/disadvantages of language communication?

Grounding

- What does it mean for something to be grounded?

Multi-Agent

- What does it mean for a system to be “multi-agent”?

Your goals

- What do you want to get from this course?

Logistics

- Assignments: Canvas
- Everything else: Course website
- By next class:
  - Fill out the presentation interest form
  - Note: the first presentation is 1 week from today, and will be graded accordingly.
- Class after that
  - I have gotten 2,000 GPU hours on TACC, but you will need a TACC account. If you don't have one already, you should make one
  - Other GPU resources (free for students):
    - Google Colab, research TPU credits
    - AWS and Azure both give a number of free credits when you join

Questions?