Ling 4480: Computational Linguistics Research Methods (CLRM)

- **The "live" course schedule can be found at this link**
- Voting for workshop topics can be done at this link

Course Number: LING-4480 (Fall, 2025)
Lecture: Tuesday, 2:00 - 4:30
Location: North GW03

Instructor: Ethan Wilcox
Office: Poulton 248 (Office Hours: Thursday, 4:00 - 5:00 pm)
Email: ethan.wilcox@georgetown.edu
Website: https://wilcoxeg.github.io/

Summary: Computational Linguistics is a fast-growing and fast-moving field. This course is intended to give advanced graduate-level students practice conducting original research in computational linguistics and to enhance their research and communication skills. It will serve as a platform for students to pursue an independent research project with guidance and oversight from faculty and peers. Students will be expected to bring their own pre-existing research topics/questions to the class. Over the course of the semester, they will develop the project with the goal of writing an ACL-style conference proceedings paper. In addition to hands-on research, this class will provide a venue for students to learn CL-related skills that often fall through the cracks of other, content-focused courses. Possible workshop topics include data annotation, LaTeX, and working with pretrained language models, as well as communication skills such as poster and slide design. As a hands-on course whose content changes based on the instructor and students, this course can be repeated for credit.

Learning Outcomes:

- Students will practice essential skills for the computational linguistics research cycle, including both technical skills and presentational skills.
- Students will gain experience conducting independent research, including developing research questions, implementing necessary methods, and sharing work with peers.

Course Structure: The course meets for 2½ hours each week. Sessions will be organized along the following structure:

- **Research Check-In (15 minutes):** Each student gives a (brief) weekly research update: what did you accomplish last week, what do you hope to accomplish during the week ahead? What are your current blockers? Is there anyone in the class that can help you overcome these?
- **Research Presentation (~ 1 hour)**: Presenter gives a research update, followed by critique and feedback. Each student is responsible for two presentations, one during the first half of the semester, one during the second half:
    - *First-half Presentations:* A brief (~ 15-minute) presentation that focuses on your research questions, associated theories, and related literature. Followed by a longer (~ 45-minute) presentation of one key paper related to your research topic.

    - *Second-half Presentations:* A longer (~ 45-minute) presentation that recaps the research question/hypothesis, and then focuses on methods, experimental design, results, statistical modeling, and interpretation. Followed by 15 minutes of discussion.
- **15-minute Break** (important to maintain mental clarity!)
- **Workshop / Guest Speakers (1 hour):** Hands-on practice with a CL-related skill, or a presentation from researchers in the field.

Course Schedule (To be filled in following First Session)

Week	Assignment	Module	Course Schedule
Week 1 (9/9)			* Course overview * Discussion of AI policy * Workshop: Designing a research question
Week 2 (9/16)			* Research Presentation: * Workshop:
Week 3 (9/23; Rosh Hashanah)	Research Proposal (due midnight)		* Research Presentation: * Workshop:
Week 4 (9/30)			* Research Presentation: * Workshop:
Week 5 (10/7)			* Research Presentation: * Workshop:
Week 6 (10/14)			* Research Presentation: * Workshop:
Week 7 (10/21)			* Research Presentation: * Workshop:
Week 8 (10/28)			* Research Presentation: * Workshop:
Week 9 (11/4)			* Research Presentation: * Workshop:
Week 10 (11/11)			* Research Presentation: * Workshop:

Week 11 (11/18)			* Research Presentation: * Workshop:
Week 12 (11/25)			* Research Presentation: * Workshop:
Week 13 (12/2)	Paper Introduction (due in class)		* Research Presentation: * Workshop: Paper introduction workshop
Week 14 (12/9)	Scientific Figures (2x; due in class)		* Workshop: Scientific Figures * Concluding Activity

Assignments:

- **Research proposal / Pre-registration:** While students are expected to bring their own research topic, a short 1-page research proposal is due in the second week. The proposal will roughly follow the design of a preregistration, and is intended to guarantee that students have a clear research direction. The research proposal should include the following (ideally as separate sections):
    - Clear outline of the research question: What hypotheses will this research attempt to answer?
    - Discussion of the broader picture: What theories will this research support or (attempt to) invalidate? What limitations in our current knowledge/methods will this research overcome?
    - Previous literature: What prior works are most relevant to the current research project?
    - Methods: What methods will be undertaken to answer the research question? Specifically:
        - What datasets and/or models will be used, if any?
        - What statistical tests will be conducted, if any?
- Helpful links for this assignment:
    - Designing a Research Question (RQ)
        - Formulation of Research Question – Stepwise Approach by Ratan et al., JIAPS
        - Write the Paper First, by Jason Eisner, JHU
    - Preregistration Design:
        - Center for Open Science webpage on preregistration, Center for Open Science
        - Open Science Foundation Preregistration Template, Center for Open Science

- **Research Presentation:** Throughout the semester students will present two research presentations. These should be informal, but detailed updates on your research project and related literature as you make progress throughout the semester. Research presentations should be presented via a slide deck, or similar. Each research presentation should have a slightly different focus:

    - *Research Presentation 1:* Brief (~ 15-minute) discussion of the "roadmap" or plan for your research project, including:
        - Detailed presentation of the research question
        - Discussion of the "so what?" – How does this topic fit into the broader research landscape? Why do you think it is important?
        - Literature review: Brief presentation of related papers. What are their limitations?

      This is followed by a longer (~ 45-minute) presentation of one (or more) related papers. Please let us know what related papers you plan to discuss approximately a week in advance, so others can have a chance to look over them.
    - *Research Presentation 2:* Discussion of methods and results
        - Presentation of experimental design: What are you measuring? What are your different experimental conditions? What are the potential confounds?
        - Detailed presentation of the methods, including datasets and models used. Discuss the tradeoffs associated with these choices
        - Detailed presentation of statistical tests (if applicable)
        - Presentation of initial results (if applicable)

  The goal of the research presentations is to give you a rough draft of material, which you can turn into your final paper at the end of the course.

- **Paper Workshop:** Students will write an "introduction" section of their final paper. In class, we will read each introduction and offer feedback. The goal of this assignment is to get you started writing your research paper well in advance of the final deadline. Introductions should (roughly) follow the following 5-paragraph format:
    - Paragraph 1: "Set the stage" – What is the broad topic of the paper? Why is it important? What is the general state of the field? Why should readers *care* about this paper?
    - Paragraph 2: "The problem statement" – What are the limitations that motivate this research question? What are the gaps in our current understanding? How has previous research fallen short?
    - Paragraph 3: "Our approach" – How does the current research project address the problem statement in paragraph 2? What new methods, ideas, or techniques do we use to overcome the current limitations?
    - Paragraph 4: "Methods" – Describe the methods in more detail. How many studies did you run? What datasets do you use? What statistical analyses were conducted? (This paragraph may be omitted for a shorter introduction.)
    - Paragraph 5: "Results" – Briefly summarize and interpret the results of your research. End by discussing the importance of these results. How have they changed our perspective on the research topic? What theories have they validated, or contradicted? Why should readers care?
- Useful links for this assignment:
    - Writing an Introduction to a Scientific Paper, by Andrew Anderson (I think this is an excellent piece!)

- **Figure Workshop:** Students will bring two scientific figures (plus associated captions) from their research project to class. In class, we will critique each figure. The goal of this assignment is to

get you working on your results section, well in advance of the final deadline. Figures should be interpretable alone, without the immediate context of the paper, and include the following:
    - Visual display of data that reports the result of the experiment
    - Well-labeled axes and legends
    - A caption that briefly explains the figure
- Useful links for this assignment:
    - The Visual Display of Quantitative Information by Edward Tufte
    - Designing Effective Scientific Figures, course by Aiora Zabala

- **Final paper:** Students will write an 8-page ACL-style final paper on their project. It's OK if the paper reports results that are still in progress; sometimes research takes more than one semester's worth of work! The goal is to practice writing in the format and style of ACL papers.
    - Students should use the ACL style files (linked below) and write the paper in Overleaf
    - The paper should include (at least) the following sections: Abstract, Introduction, Methods, Results, Discussion, Limitations, and Ethics Statement. The results section should include figures, tables, or other graphics necessary for reporting the results of the experiments.
- Helpful links for this assignment
    - Nuts and Bolts:
        - ACL Rolling Review call for papers
        - Association for Computational Linguistics paper format guidelines and style files
    - Writing Advice
        - Tips for Writing NLP Papers by Vered Schwartz
        - Style Points for Scientific Writing, University of Washington

Participation and Attendance: The grade for participation and attendance will be based on the following criteria:

- Check-ins: Students will give a brief (2-minute) oral update about their research at the beginning of our weekly meeting. Mention any roadblocks you are encountering.
- Feedback for Research Presentations: Share constructive feedback after others' research presentations. What were the strengths of the presentation or idea? What parts did you not understand? Where do you see room for improvement?
- Participation in workshops: Engage in group discussion as we learn skills and discuss issues in the field.
- Attendance: Students are expected to attend class in person, except for personal or medical emergencies, or religious commitments.

Prerequisites: This course is intended for advanced graduate students who are engaged in original research. As such, we expect computer programming at the level of "Advanced Python" (i.e., the second course in our programming sequence) or better.

Group Work: Working in teams is an integral part of doing research in computational linguistics. However, the goal of this course is to strengthen *individual* skillsets. Therefore, students will conduct

their research projects individually, and not in groups. If two students wish to work on similar or related topics, please discuss the matter with Ethan.

AI Policy: The main goal of this class is to develop students' independent research skills. Effective use of AI is one such skill. However, AI can also be used as a crutch. Especially when it comes to writing, overuse of AI can often obfuscate and lead to confusion. Therefore, while AI tools are not banned, I am going to *limit* their use to intermediary stages in the writing and coding process.

- You are expected to write the first draft *and* the last draft of any assignment submitted. AI tools can be used to refine the flow of a paragraph, or make the English text sound more native-speaker. In general, use AI tools to rewrite sentences, not whole paragraphs.
- You are expected to design the logic and debug any code or computational experiments. AI tools can be used, for example, to implement functions.
- I do encourage you to use AI in a dialectical manner to brainstorm and refine ideas.
- I do encourage you to use AI tools such as perplexity.ai and elicit.com to perform literature reviews.

For any assignment, if you use AI tools, I will ask that you submit a **supplementary document** that details where and how you used them. You do not need to report all sessions or prompts, but please provide a high-level overview of the role that AI played in the completion of the assignment.

Grade Breakdown

Final paper	30%
Paper Introduction Workshop	10%
Scientific Visualization Workshop	10%
Research Presentation	30%
Participation and Attendance	20%

List of Possible Workshop Topics

- I will solicit suggestions for additional topics on the first day of class
- We'll vote on topics and make a final list after the first class session

Topic	**Description / Possible Activity**	
Working with LLMs and Huggingface	* Discussion of use cases for LLMs in CLI research * How to choose an appropriate LLM * Discussion of small vs. large models	
Running Studies with Crowdworkers	* How to host a study on Prolific / MTurk * Challenges of working with human annotation / psycholinguistic data * IRB approval and ethics: what you need to know	
Working with Equations / Advanced Typesetting	* How to present proofs, equations, and formal mathematical arguments	

	* Advanced typesetting with LaTeX (macroing, etc.)	
Experimental Design and Psycholinguistics	* Factorial experimental design * Introduction to psycholinguistic methods (eye-tracking; likert scales; self-paced reading, reaction times)	
Writing Reviews and Rebuttals	* How to write an effective review * How to write an effective rebuttal to a review	
Command Line and Working on a Cluster	* Working on the command line, working with the cluster * Deploying web servers on the cluster * Introduction to cluster computing * How to deploy jobs on a cluster	
Paper Reading	* Strategies for reading papers effectively * Strategies for sourcing papers and keeping up-to-date on research in a particular area	
GIT	* Discussing git workflows	
Publishing in CLI	* Discussion of ACL conferences, how to select the right venue for your work * Submitting to a journal and the review process	
Ethics in Science and CLI	* Ethical considerations of working with LLMs and language technologies * Conducting Ethical Science: reporting bias, reproducibility, p-hacking, etc.	
Doing science in an age of LLMs / Scientific Productivity	* Using LLMs for writing and literature review * Using LLMs for coding and other aspects of research * Other scientific productivity tools (e.g., Zotero) * Role of LLMs in our scientific theories	
Collaborating across Disciplines	* Discussion of CLI + {Law, Medicine, Public Policy} * Panel from Massive Data Institute	
Designing Slides	* Best practices for designing and presenting slides	
Designing Posters	* Best practices for designing and presenting posters	
Public Speaking 101	* Basic public speaking techniques (voice, breath, body) * Presenting at scientific conferences	