Course Information

Course Number:	CSCE 689
Course Title:	Special Topics in Natural Language Processing for Science
Section:	603
Time:	TR 3:55PM-5:10PM
Location:	HRBB 126
Credit Hours:	3

Instructor Details

Instructor:	Yu Zhang
Office:	PETR 222
Phone:	217-954-3457
E-Mail:	yuzhang@tamu.edu
Office Hours:	R 2PM-3PM at my office

Course Description

A special topics course focusing on applying state-of-the-art natural language processing techniques (e.g., large language models) to the science domain, including retrieving and understanding relevant scientific literature, extracting scientific knowledge, building connections between text and scientific data of other modalities (e.g., molecules and proteins), generating novel scientific ideas and hypotheses, as well as automatically providing constructive feedback on research outcomes.

Course Prerequisites

Graduate classification

Special Course Designation

N/A

Course Learning Outcomes

At the end of study, students will be able to:

- List standard architectures and techniques for large language model pre-training in the science domain.
- Describe important natural language processing tasks in different scientific fields.
- Articulate the need of advanced natural language processing techniques for augmenting and accelerating scientific discovery.

- Apply scientific large language models to real-world scientific applications.
- Evaluate the strengths and weaknesses of representative scientific natural language processing techniques.
- Conduct literature review, present state-of-the-art research, and propose new ideas.

Textbook and/or Resource Materials

Required Textbooks

This course does not mandate any textbook. The lecture slides/videos and other materials provided by the instructor will be sufficient, serving as the primary reference. In addition, the students are recommended to refer to the following textbooks:

- ***Speech and Language Processing***, Dan Jurafsky and James H. Martin, 2024.
  - Available online: https://web.stanford.edu/~jurafsky/slp3/
- ***Dive into Deep Learning***, Aston Zhang, Zack C. Lipton, Mu Li, and Alex J. Smola, 2023.
  - Available online: https://d2l.ai/

Grading Policy

- *Participation* *10%*
- *Literature Review* *10%*
- *Paper Presentation* *20%*
- *Project* *60%*
  - *Proposal* *5%*
  - *Mid-term Spotlight Presentation* *5%*
  - *Mid-term Report* *10%*
  - *Final Presentation* *15%*
  - *Final Report* *25%*
- *There will be no mid-term or final exam.*
- *The grading scale will be:*
  - A = 90-100
  - B = 80-89
  - C = 70-79
  - D = 60-69
  - F = <60

Description of Graded Work

- **Participation (10%)**: Students are expected to submit pre-lecture questions, attend lectures, and engage in discussions. Notes on pre-lecture questions: the task is to read the papers to be introduced in each student/guest lecture and submit a question they come up with. The deadline is one day before the lecture (e.g., For Thursday lectures, they need to submit the question by Wednesday 11:59PM).

- **Literature Review (10%)**: Students are expected to submit a review for a paper introduced in the lectures. The review should follow the conference peer review format and include a paper summary, strengths, weaknesses, questions to the authors, a soundness score, and a novelty score.
- **Paper Presentation (20%)**: Students will present 3 papers they sign up for. The primary objective is to impart knowledge to the rest of the class. The presentation will be followed by a question-and-answer session with the audience. Students are required to send the slides to the instructor 2 days before their presentation (e.g., For Thursday lectures, they need to submit the slides by Tuesday 11:59PM). Rubrics for this part:
    - **Slides (5%)**: Submit complete slides before the deadline.
    - **Presentation Completeness (5%)**: Adequately cover the core concepts and insights contained in the paper.
    - **Presentation Clarity (5%)**: Explain own understandings of the paper in the presentation (e.g., raise a new example when introducing some concepts, list some limitations not mentioned in the paper).
    - **Question and Answering (5%)**: Effectively answer the questions raised by the audience.
- **Project (60%)**: Students are required to work in a team of 2 or 3 to complete a research project (either a survey or a hands-on project) related to any topic introduced in this course, present their results, and submit project reports. Rubrics for a hands-on project are as follows. Rubrics for a survey should be compared and conducted accordingly.
    - **Project Proposal (5%)**: List the team members, motivation, task definition, potential datasets, baselines, and expected outcomes of the project. The proposal should be 2-3 pages (ACL 2024 template, excluding references).
    - **Mid-term Spotlight Presentation (5%)**: Introduce task background and formal definition. Present potential solutions and preliminary results using figures and tables. The presentation is graded based on completeness and clarity.
    - **Mid-term Report (10%)**: The report should include task definition, related work, method, preliminary results, unfinished parts, and a timeline to finish them. The report should be 4-6 pages (ACL 2024 template, excluding references). The report is graded based on completeness, clarity, and efforts.
    - **Final Presentation (15%)**: Introduce task background, formal definition, related work and their limitations, proposed solution, datasets, complete experimental results, conclusions, and future work. The presentation is graded based on completeness and clarity.
    - **Final Report (25%)**: The report should follow the format of a conference paper submission, including introduction, related work, preliminaries, method, experiments, conclusions, and future work. The report should be 7-8 pages (ACL 2024 template, excluding references, an appendix is allowed). The report is graded based on novelty, soundness, completeness, clarity, and efforts.

Late Work Policy

No late work will be accepted.

Work submitted by a student as makeup work for an excused absence is not considered late work and is exempted from the late work policy (Student Rule 7). The student is responsible for informing the instructor in a timely manner about excused absences. The instructor will then work with the student to catch up on their submissions, including missed group class participation.

Course Schedule

Week	Topics
1	Introduction; Scientific Large Language Models
2	Scientific Large Language Models
3	Understanding Scientific Literature: Citation Prediction and Literature Retrieval
4	Understanding Scientific Literature: Question Answering, Knowledge Extraction, and Paper Classification
5	Scientific Vision-Language Models: Bioimaging
6	Scientific Vision-Language Models: Geometry and Geoscience
7	Protein Language Models
8	Molecule Language Models
9	Urban Language Models; Mid-term Project Presentations
10	Language Models with Academic Graphs; Table Language Models
11	Large Language Models for Scientific Discovery: Idea and Content Generation
12	Large Language Models for Scientific Discovery: Reviewing
13	Large Language Models for Scientific Discovery: Execution and Revision
14	Final Project Presentations
15	Final Project Presentations

Important Dates (Dates are specific to Spring 2025)

- Pre-Lecture Questions: One day before the lecture
- Literature Review: Mar 7
- Presentation Slides: Two days before the presentation
- Project
    - Project Proposal: Feb 23
    - Mid-term Report: Mar 23
    - Final Report: May 4

University Policies

Attendance Policy

The university views class attendance and participation as an individual student responsibility. Students are expected to attend class and to complete all assignments.

Please refer to Student Rule 7 in its entirety for information about excused absences, including definitions, and related documentation and timelines.

Makeup Work Policy

Students will be excused from attending class on the day of a graded activity or when attendance contributes to a student's grade, for the reasons stated in Student Rule 7, or other reason deemed appropriate by the instructor.

Please refer to Student Rule 7 in its entirety for information about makeup work, including definitions, and related documentation and timelines.

Absences related to Title IX of the Education Amendments of 1972 may necessitate a period of more than 30 days for make-up work, and the timeframe for make-up work should be agreed upon by the student and instructor" (Student Rule 7, Section 7.4.1).

"The instructor is under no obligation to provide an opportunity for the student to make up work missed because of an unexcused absence" (Student Rule 7, Section 7.4.2).

Students who request an excused absence are expected to uphold the Aggie Honor Code and Student Conduct Code. (See Student Rule 24.)

Academic Integrity Statement and Policy

"An Aggie does not lie, cheat or steal, or tolerate those who do."

"Texas A&M University students are responsible for authenticating all work submitted to an instructor. If asked, students must be able to produce proof that the item submitted is indeed the work of that student. Students must keep appropriate records at all times. The inability to authenticate one's work, should the instructor request it, may be sufficient grounds to initiate an academic misconduct case" (Section 20.1.2.3, Student Rule 20).

You can learn more about the Aggie Honor System Office Rules and Procedures, academic integrity, and your rights and responsibilities at aggiehonor.tamu.edu.

Americans with Disabilities Act (ADA) Policy

Texas A&M University is committed to providing equitable access to learning opportunities for all students. If you experience barriers to your education due to a disability or think you may have a disability, please contact the Disability Resources office on your campus (resources listed below) Disabilities may include, but are not limited to attentional, learning, mental health, sensory, physical, or

chronic health conditions. All students are encouraged to discuss their disability related needs with Disability Resources and their instructors as soon as possible.

Disability Resources is located in the Student Services Building or at (979) 845-1637 or visit disability.tamu.edu.

Title IX and Statement on Limits to Confidentiality

Texas A&M University is committed to fostering a learning environment that is safe and productive for all. University policies and federal and state laws prohibit gender-based discrimination and sexual harassment, including sexual assault, sexual exploitation, domestic violence, dating violence, and stalking.

With the exception of some medical and mental health providers, all university employees (including full and part-time faculty, staff, paid graduate assistants, student workers, etc.) are Mandatory Reporters and must report to the Title IX Office if the employee experiences, observes, or becomes aware of an incident that meets the following conditions (see University Rule 08.01.01.M1):

- The incident is reasonably believed to be discrimination or harassment.
- The incident is alleged to have been committed by or against a person who, at the time of the incident, was (1) a student enrolled at the University or (2) an employee of the University.

Mandatory Reporters must file a report regardless of how the information comes to their attention – including but not limited to face-to-face conversations, a written class assignment or paper, class discussion, email, text, or social media post. Although Mandatory Reporters must file a report, in most instances, a person who is subjected to the alleged conduct will be able to control how the report is handled, including whether or not to pursue a formal investigation. The University's goal is to make sure you are aware of the range of options available to you and to ensure access to the resources you need.

Students wishing to discuss concerns in a confidential setting are encouraged to make an appointment with Counseling and Psychological Services (CAPS).

Students can learn more about filing a report, accessing supportive resources, and navigating the Title IX investigation and resolution process on the University's Title IX webpage.

Statement on Mental Health and Wellness

Texas A&M University recognizes that mental health and wellness are critical factors that influence a student's academic success and overall wellbeing. Students are encouraged to engage in healthy self-care by utilizing available resources and services on your campus.

Students who need someone to talk to can contact Counseling & Psychological Services (CAPS) or call the TAMU Helpline (979-845-2700) from 4:00 p.m. to 8:00 a.m. weekdays and 24 hours on weekends. 24-hour emergency help is also available through the 988 Suicide & Crisis Lifeline (988) or at 988lifeline.org