CS 7170 *Seminar in Artificial Intelligence*
Multilingual Language Modeling

Time: MR 11:45-1:25
Location: WVH 366
Instructor: Terra Blevins
Office Hours: MR 1:45-2:45 (after class) in 177 Huntington, Room 2225
Canvas: TBA

Description
This seminar provides an overview of multilingual NLP in the era of generative modeling. We will discuss the fundamentals of multilingual language processing as well as recent advances (and bottlenecks) in the field due to LLMs. The course will involve lectures introducing topics in multilingual modeling, student-led paper discussions, and a final project proposal.

Prerequisites
There are no formal prerequisites for this course, but students should have some familiarity with machine learning methods and natural language processing.

While there are no language or package requirements for the computational portion of the project proposal, most research on language models is done in Python and with the Huggingface packages. [Here is a tutorial to getting started working with LLMs in the Huggingface ecosystem.](#)

Coursework and Grading
Participation + Presenting (40%): The primary focus of this course will be on paper discussions of seminal and recent works in multilingual NLP from the assigned reading list. Students will be expected to present papers (from the PDF) and lead discussions of the works throughout the semester, and to participate in these discussions when not presenting.
- **EMNLP Recap:** Students will create summary presentations of multilingual NLP @ [EMNLP 2025](#) to discuss the most recent work in the field, and present their findings to the class on 11/10.

Project Proposal (60%): The final project for this course is a *proposal*, accompanied by *initial results* and a *literature review*, for a novel multilingual NLP project with LMs; the primary requirement for this project scope is for the research questions to address multiple languages rather than just one. The final version of the proposal should also include a description of, and initial results from, a proof-of-concept for the proposed research direction. The project consists of four (4) deliverables, outlined below:
- **Proposal (5%, due 09/29)**: A one-page summary outlining the motivation, hypothesis, and approach for the student's project. This portion is analogous to the introduction to a conference paper submission.
- **Literature Review (15%, due 10/30)**: A comprehensive review of related works to the student's project proposal. This proposal can be written similarly to a paper's related work section; however, the write-up should be *complete* and *thoroughly engage with all of the included works*. This submission should consist of (1) the completed literature review and corresponding

bibliography and (2) a revised proposal that addresses the relationship between the project as proposed and existing prior work.

- **Final Report (25%, due 12/11)**: This writeup should include your original proposal, a summary of your literature review, a description of your proof of concept implementation, and a discussion of the initial results and their implications for the research direction. The final write-up should be no more than four pages in length.
- **Final Presentation (15%, due 12/11):** Students will give a presentation based on their project proposal and initial findings to the class. These presentations should take the form of a research conference talk, and students should expect to answer questions from the instructor/class about their project.

Tentative Schedule (TBD)

Topic	Date		Reading/ Assignments Due
Introduction	09/04	Overview of the syllabus and introduction to multilingual language modeling	Background Reading: chapters 10, 11, and 12 of *Jurafsky and Martin (2025)*, particularly if you have less familiarity with recent advances in NLP.
History of multilinguality in NLP	09/08		
	09/11		
Foundations of multilingual language modeling	09/15		
	09/18		
The curse of multilinguality	09/22		
	09/25		Due: Project Proposals
Architectures for Multilingual Modeling	09/29		
	10/02		
Multilingual Applications of Language Models	10/06		
	10/09		
–	10/13	No Class for Indigenous Peoples Day	
Guest Lecture	10/16		
Tokenization	10/20		

	10/23		Due: project literature review
	10/27		
Cross-lingual Transfer	10/30		
	11/03		
	11/06		
EMNLP Recap	11/10	Summary presentations of trends in multilingual NLP @ EMNLP 2025	Due: student presentations on EMNLP findings
Low- and No-Resource Language Modeling	11/13		
	11/17		
	11/20		
Multiculturalism in Language Models Multiculturalism in Language Models	11/24		
	11/27	No Class for Thanksgiving	
	12/01		
Current State of Multilingual Language Modeling	12/04		
	12/08		
Project Presentations	12/11		Due: student presentations on final project Due: final project reports

Course Policies

Attendance and Tardiness: Classes will be held in-person each week -- any deviation in the delivery of the lecture will be announced through Canvas. Attendance and active engagement during lecture is highly recommended, and presenting and participation in class each week is a key part of your learning experience and grade. *If you have circumstances that can prevent you from being in class on time, please email the instructor in advance (if possible).*

Make-up Policy: The majority of course assignments are based on in-class participation. For submitted assignments throughout the semester, students can obtain late days for legitimate reasons such as illness and family emergencies; *however, these cases need to be approved prior to the submission deadline by the instructor.* No late days can be used for the final project deliverables.

Regrade Considerations: If you are confused or concerned about feedback on course assignments,

please directly email the instructor. You must submit any requests for grading reconsideration within 7 days after the feedback was released.

Academic Integrity: Please read the [Northeastern Academic Integrity Policy](#). All students are required to adhere to this policy during the course. Note that while students are encouraged to discuss course materials, no plagiarism/copying is allowed.

This course also has a firm no AI usage policy for completing the following portions of the coursework: *reading* (do not have a model summarize or otherwise explain the paper for you), *presenting*, and *writing*. All work submitted to this course should be primarily completed by you, without the help of an AI. The following cases are exceptions to this policy:
- *Coding Support*: You may use an AI such as CoPilot to help you with the project implementation.
- *Grammar*: You may use AI support to help you edit your writing, once you have produced the first draft. The AI may help you with grammar and wording, but the content of the text should be wholly produced by you.

However, ***in any case where you use AI***, you must thoroughly check the output of the model. If you choose to use AI for these cases, you take ownership of the final product, regardless of any errors or artifacts introduced by the AI.

The first time you are found in violation of the policies in this syllabus on an assignment, you will receive a 0 for the associated work. A second violation, or a violation during the final project deliverables, will result in failing the course.

Classroom Environment: To create and preserve a classroom atmosphere that optimizes teaching and learning, all participants share a responsibility in creating a civil and non-disruptive forum for the discussion of ideas. Students are expected to conduct themselves at all times in a manner that is respectful towards all participants and does not disrupt teaching or learning. The instructor reserves the right to interrupt conversations that deviate from these expectations. Repeated unprofessional or disrespectful conduct may result in a lower grade or more severe consequences.

Title IX: Northeastern's Title IX Policy prohibits Prohibited Offenses, which are defined as sexual harassment, sexual assault, relationship or domestic violence, and stalking. The Title IX Policy applies to the entire community, including male, female, transgender students, faculty and staff.

If you or someone you know has been a survivor of a Prohibited Offense, *confidential* support and guidance can be found through **University Health and Counseling Services** staff and the **Center for Spiritual Dialogue and Service** clergy members. By law, those employees are not required to report allegations of sex or gender-based discrimination to the University.

Alleged violations can be reported non-confidentially to the Title IX Coordinator within **The Office for Gender Equity and Compliance** at titleix@northeastern.edu and/or through **NUPD** (Emergency 617.373.3333; Non-Emergency 617.373.2121). Reporting Prohibited Offenses to NUPD does **NOT** commit the victim/affected party to future legal action.

Faculty members are considered "mandatory reporters" at Northeastern University, meaning they are required to report all allegations of sex or gender-based discrimination to the Title IX Coordinator.

Please visit https://www.northeastern.edu/titleix for a complete list of reporting options and resources both on- and off-campus.

Students with Disabilities: Students who have disabilities who wish to receive academic services and/or accommodations should visit the Disability Resource Center at 20 Dodge Hall or call (617) 373-2675. If you have already done so, please provide your letter from the DRC to me early in the semester so that I can arrange those accommodations.

References:

Tyler A Chang, Catherine Arnett, Zhuowen Tu, and Benjamin K Bergen. 2024. When is multilinguality a curse? Language modeling for 250 high-and low-resource languages. In Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2024.

Daniel Jurafsky and James H. Martin. 2025. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models, 3rd edition. Online manuscript released January 12, 2025. https://web.stanford.edu/~jurafsky/slp3.