

CS 6120: Graduate Natural Language Processing, Fall 2026

Instructor: Prof. Felix Muzny¹ (/ˈfi.lɪks/ /ˈmju.z.ni/ or "Fee-licks Muse-knee"; pronouns: they/them and he/him²)

Contact: f.muzny@northeastern.edu

Office: Meserve 307A

TAs:

Chaitanya Agarwal (/ˈtʃeɪ.tən.jɑ/ or "Chay-tun-yah", pronouns: he/him)

Shrivarshini Narayanan ("Sh-ree-var-shee-nee"; pronouns: she/her)

Utkarsh Saraogi ("Oot-kursh", pronouns: he/him)

Contacting course staff: we prefer that you pose all general questions on the [course piazza](#), find the sign-up under the "Piazza" item in the General module on Canvas

Gradescope: course access code JN3DBW

Contacting Prof. Felix: reach out via email or before/after class. Please do not message them via teams.

Credit: 4 credits, Tuesdays & Fridays, 9:50 -11:30am, Hurtig Hall 129

Office Hours: see Canvas

Lecture Format: Participating in synchronous course content will be important to your overall learning and is expected (and in some cases graded). If you need to attend lecture remotely on a certain day, send Felix an email at least 1 hour before lecture begins—they will send you a zoom link. If you are attending remotely, expect to be actively participating in all course content. Not all lectures will have a remote option—this will be noted on the course schedule and/or announced in class.

As an on-ground class, it is expected by the university that the primary mode of instruction is in-person, on-ground. Do what you need to do to be fed/hydrated/prepared to be present in lecture.

There are a few items that I'd like to emphasize:

- *Life happens, **communicate** as proactively as you can and we will work together to make a plan.*
 - *Practicing proactive communication is a valuable skill that will help you in your post-university life.*
 - *If you are sick in any way, do not come to class in person, even if you have received a negative covid/flu/other test. If you are well enough to attend class remotely, this is the best alternative option.*
 - *I will do the same, as will our TAs.*
-

¹ Call them "Felix", "Professor Muzny", or "Professor Felix"

² This means that they are happy with either they/them or he/him or a mixture of both sets of pronouns

Some Pedagogical Notes

I frequently find that a small amount of meta-learning is helpful to student success in courses. I am always happy to share with you *why* an activity/assignment/lecture is designed the way that it is.

Here are some quick facts that we know about *learning*:

1. Learning happens by doing, and, importantly, revising previous misconceptions.
2. Learning takes two important ingredients: *time* and *repetition*.
3. Group work is most effective when:
 - a. all members of the group have similar experience with the topic *and discuss their understanding of the problem together* OR
 - b. with different levels of experience, if the group member(s) with more experience take on a "peer educator" type role *without* "talking down" to their group members; that is, the group as a whole becomes invested in achieving equal understanding of the topic.
4. No one can actually multi-task.
 - a. *When you choose to do other work in class, you are choosing not to learn the current material. Making the choice to focus just on class is difficult in today's age of instant communication and over-stimulation; it's difficult for professors when they are in meetings and have looming deadlines as well! You might find that making this choice in one course makes focusing in all parts of your life easier.*

Here are some quick facts we know about *retaining information*:

1. Writing information down yourself is more effective than just reading or just listening.
2. Re-wording information in your own words is more effective than copying pre-written words.
3. Writing information by hand is more effective than typing.

Please talk to Felix if you want some help with any of these or if you want more resources to learn about learning!

Course Overview

NLP is about getting computers to perform useful and interesting tasks involving spoken and written human language. NLP is sometimes referred to as Computational Linguistics to emphasize the fact that it involves the combination of CS methods with research insights from Linguistics (the study of human language). Practical applications of NLP include question answering, machine translation, information extraction, and interactive dialog systems (both written and spoken). Modern NLP systems rely heavily on methods involving probability, linear algebra, and calculus --- often in combination with machine learning methods.

We'll be exploring both applications and the computational methods behind them. You should be prepared to get your hands dirty in terms of the math, programming, and data that comprise the behind the scenes components of NLP systems.

Expect a heavy emphasis on data, how systems are deployed in the real world, and engaging in critical thought about provided reading materials.

NLP is not *just* generative AI, and you can expect a significant fraction of this course to inspect other aspects of NLP (even though we'll cover generative AI models like ChatGPT, GPT-3, GPT-4, etc).

Course Goals

1. Develop an understanding of the general problems that people who work on NLP study and the strategies they use to solve them.
2. Understand the role of data, machine learning, and neural networks in NLP systems.
3. Understand the ethical considerations and potentials for bias in NLP systems.
4. Be able to implement models to solve some "standard" NLP problems.
5. Be able to formulate potential starting points given a new problem with NLP elements.
6. Understand some of the motivating linguistic phenomena that make NLP problems hard and why these can be hard phenomena for computers to approach.

Topics

- Language models, large and small
- Probabilistic (non-neural) methods & models
- Issues of ethics & bias in NLP
- Data sets and characteristics
- Words, word counting, lexicons
- Text classification with language models
- Text classification with single layer neural networks
- Vector semantics & word embeddings
- Part-of-speech tagging
- Viterbi algorithm & dynamic programming
- Machine translation
- Speech-to-text and text-to-speech*
- Information extraction*
- Question answering systems*
- Dependency parsing*

* if time allows

Textbook & Course Configuration

1. We'll be using draft chapters from the 3rd Edition of *Speech and Language Processing* by Dan Jurafsky and James H. Martin. You don't need to buy the current edition, draft pdfs of the new chapters are available from [the textbook website](#). You can also download (and print, if you desire) the entire book from the website. We will also link a pdf of this text from the course website.
 - a. We will be using the draft version from **August 19th, 2026**
2. We will supplement this text occasionally with readings from:
 - a. [Eisenstein, Jacob. *Introduction to Natural Language Processing*. MIT Press, 2019.](#)
 - b. [Kohn, Philipp. "Neural Machine Translation." arXiv preprint arXiv:1709.07809, 2017.](#)
 - c. (and others, to be linked from the course website)

Websites & Technology

Make sure that you have access to all of the following websites and software:

- **Canvas**: posted course materials and links to homework submissions.
- **Gradescope**: where you will submit most coursework.
- **Piazza**: our course discussion forum. This is where we will discuss relevant topics and answer your homework and content questions that come up outside of class. If you send us a content question via email or Teams, we'll ask you to post it to Piazza instead!

- **Python 3:** We'll be primarily coding with python 3. Many assignments will ask that you are familiar with object-oriented programming in python, running unit tests, and running both .py files and jupyter notebooks.
 - For help with python environment management: post your problem (screenshots are always encouraged) on piazza or come to office hours.
- **Jupyter Notebooks:** Many of our coding activities will be distributed as Jupyter Notebooks. You can install jupyter notebooks several ways, such as with [VS Code](#), [Anaconda](#), or via the [command line](#).
- **IDEs:** You can develop your code using whatever your preferred IDE is.
 - If you have less experience working with python and **both** Jupyter Notebooks and .py files, we **highly** encourage you to make time to come to office hours in the first few weeks of the course.

Classroom Environment & Expectations

- **Preparation:** *this course will involve significant reading.* It is the expectation that you do reading by the lecture it is associated with. This course will be a great opportunity for those of you who are interested in NLP & research to start flexing those muscles, and the best way for us to go down those paths is for you to develop a solid foundation of reading research papers. *You are not expected to understand everything you read.*
- **Attendance:** You are expected to attend lecture *synchronously* whenever possible. **You are expected to attend in person whenever this is a reasonable choice.** We will be doing interactive activities during lecture as well as covering the material necessary for you to complete your homework and quizzes. Past students have reported the in class materials to be very valuable to their success in this course.
- **Classroom environment:** It is unusually common in Computer Science classes for some students to ask questions that are not really questions so much as opportunities to demonstrate knowledge of vocabulary or facts beyond the topic at hand. This can have a discouraging effect on other students who are not familiar with those terms, causing them to worry that they are less prepared to do well in the class (this is rarely the case—knowing terms outside the scope of the course is not a good predictor of success). If you find yourself wanting to make such a question or comment, please come talk to me about the topic after class or during office hours—I'm always happy to discuss tangentially related topics at those times! Please be particularly cognizant of using acronyms that we haven't yet talked about as a group.
- **Accommodation letters:** If you have an accommodation letter, please email it to me at your earliest convenience so that I can make sure this class is meeting your needs.
- **Name and pronouns:** If your name and pronouns are not in alignment with those listed on our class roster, please let me know either in person or via email so that I can ensure you are correctly addressed.
 - to add, change, or update your pronouns in Canvas, go to "Account" > "Profile" > "Edit Profile", then add, change, or update your pronouns and display name.
 - to change or update your name here at Northeastern as a whole, find [instructions with the registrar here](#).
- **Class expenses:** If obtaining any material for use in our class presents a financial hardship for you, please let me know and I will work with you to locate the resources that you need for this class. You are **not expected to pay for compute time or LLM usage, including Claude credits** for any components of this course. Please talk to me if you find yourself in a situation where you are thinking about doing this.
- **Feedback:** Please don't hesitate to reach out to me if any aspect of this course or class community could be improved.

- **Illness:** If you are ill in any way, you are expected to attend class remotely or make up the material asynchronously on your own and by attending office hours.

Late Policy

All homework should be turned in on time whenever possible. All homework may be turned in up to 24 hours late for a 5% penalty. For example, if homework is due on Thursday at 9pm, it may be turned in as late as Friday at 9pm.

Readings cannot be turned in late.

Extensions

Extensions beyond the regular late policy will be given based on proactive communication with Felix.

Whenever possible, this should occur at least 24 hours before the posted deadline. The sooner that you reach out, the easier this will be.

Email Felix (f.muzny@northeastern.edu) with the following information:

- 1) Which assignment are you requesting an extension on & why you are requesting an extension.
- 2) When are you requesting the extension until.
- 3) What is your plan for how this extension will impact the due dates for the other assignments in this course.

Please don't write Felix an essay. This extension policy is based on our mutual understanding that life happens, we're all doing our best, and the best muscle to practice using for the real world is proactive communication. If a situation arises that makes it impossible to reach out 24 hours before the deadline, don't panic—send an email as soon as you can and we'll discuss your options together.

Fall 2026 AI Collaboration Policy

1. How should LLMs be allowed to be used in this course? Justify your answers.

Students generally felt that LLMs should be allowed as a learning and support tool, especially when they help students better understand course material ***without replacing*** their own work.

Students should ***always*** spend 10 - 20 minutes brainstorming about, designing an approach to, and working on problems on their own before using an LLM to aide them.

LLMs can be used to:

- clarify difficult concepts or explain them in a different way,
- ask follow-up questions,
- debug code and understand errors,
- check work after the student has attempted it,
- improve grammar, clarity, or formatting,
- and help with smaller implementation or coding tasks.

The main expectation is that students should still understand and be able to explain anything they submit. LLM outputs should also be checked rather than automatically treated as correct.

If an LLM meaningfully contributes to submitted work, students should be transparent about how it was used.

2. How should LLMs not be allowed to be used in this course? Justify your answers.

Students generally agreed that LLMs should not be used to replace the main learning or reasoning that an assignment is designed to assess.

LLMs should not be used to:

- generate complete homework answers,
- write an entire coding assignment,
- do most of the reasoning for a problem,
- copy an assignment question into an LLM and submit the response,
- skip required course material by relying only on AI-generated summaries,
- complete exams or assessments unless explicitly allowed,
- or generate whole projects ideas, define project scope, or the majority of project details

Students were more open to limited coding help, debugging, or small suggestions, as long as the student still develops and understands the overall solution.

Overall, the common idea was that LLMs can help with the work, but they should not do the work for you.

3. On a scale from 1 - least permissive (no large language models) to 5 - most permissive (large language models are always okay), what level of permissiveness does the class want for this course?

3.5 - 4

4. Translate your answer from question #3 into one or two sentences.

Students generally want LLMs to be available for learning, clarification, brainstorming, debugging, checking work, and limited coding assistance. Since this is an NLP course, students felt that LLM use should not be completely restricted, but it also should not replace the student's own reasoning or be used to generate full assignment solutions.

5. What do you think should happen if you use an LLM in a way that does not meet this policy?

Students generally preferred a progressive approach to violations, with some groups wanting very lenient approaches and some groups wanting very harsh approaches.

(See the general Collaboration Policy for text on consequences)

Collaboration Policy

The work that you turn in should be your own. We encourage you to collaborate with your classmates, but remember that collaboration does not mean "completing an assignment together".

Here are three big-picture points to remember when collaborating with your classmates in non-group work settings:

- **Strategies:** You may talk with your classmates about *general strategies* but you may not talk about *specific solutions*.
- **Explaining concepts:** You may talk with your classmates about how certain techniques work *in general* but not how to write any part (or sub-part) of the solution needed for the homework.
- **A good rule of thumb:** don't show your assignments to other people; don't look at other people's assignments (this makes it *very* hard to come up with your own solution afterwards); don't write code together unless explicitly stated. This includes working through solutions on whiteboards and talking through specific solutions verbally.

You are expected to use the internet as a place for online resources, such as documentation, not as a place to get solutions to your assignments.

The finer-grained details:

- **Do not search for a solution online:** You may not actively search for a solution to the problem from the internet. This includes posting to sources like StackExchange, Reddit, Chegg, etc.
 - **StackExchange Clarification:** Searching for basic techniques in python is fine. If you want to post and ask "How do convert a float to an integer" that's fine. What you **cannot** do is post things like "Here's the function my prof gave me to write. I need to convert this temperature in celcius to farenheit".
- **Plagiarism:** assignments and code should be your own. You should not need to consult sources beyond the class notes, posted lecture notes, examples, resources, and python and its associated libraries' documentation. Make sure that you understand how this relates to the AI collaboration policy for this course.
 - We will practice properly documenting other resources you consult throughout this class.
- **Tutors:** you should always consult the course instructional staff if you need extra help. They are here specifically to help you! You should never have anyone else write code for you. This includes tutors, friends, strangers, friends of friends, or anyone who is not you (even friends who have taken this course previously!).
- **When in doubt, ask:** If you have doubts about this policy or would like to discuss specific cases, please ask Prof. Felix.

For Collaboration Policy violations, students will receive 1 warning and be required to resubmit the work in question, and the instructor may require the student complete similar but different work. Further violations will result in 0s on the assignments in question.

Violations on exams will result in a 0 with no warning.

More than three violations (inclusive of a warning, across both AI Collaboration Policy and general Collaboration Policy) will result in an F in the course and will be reported to the appropriate university-level committee. For graduate students, consequences of violating collaboration policies vary but often result in being barred from being eligible to TA for the duration of your time at Northeastern in addition to other consequences.

The university's academic integrity policy discusses actions regarded as violations and consequences for undergraduate students in particular. <http://www.northeastern.edu/osccr/academic-integrity>

Grading & Assignments

If you enrolled in this course after any deadlines have passed **and** you contact Prof. Felix by the end of the

next lecture day, we will work to adjust deadlines as needed.

Category (number)	Due Dates & Notes	Weight
Homeworks	HWs are due on Fridays. Expect overlap between homework material, in class questions, and exam material.	15%
Readings	You will have 1 - 3 short questions related to reading. Due before class on days assigned.	5%
Presentation	You will have one research paper presentation during the term. These will be in <i>randomly assigned</i> groups of 2-3 people. You will read the paper in depth and prepare a 10 minute in-person presentation summarizing it. Each group member must contribute equally.	10%
Final Project	Final projects may be completed in groups of 3 - 4 people (your choice of partners). Expect there to be several intermediate deadlines for the Final Project. Posted instructions will include further details about project evaluation.	25%
Exams (2)	There will be one midterm (20%) and one final exam (25%) for this course—on Oct. 16th, 2026 and Dec. 15th, 2026. <i>Do not book travel before the final exam.</i>	45%

Final grades will be based on the following scale. Decimals will be rounded to the nearest whole number.

Letter	Range (percent)
A	95 - 100
A-	90 - 94
B+	87 - 89
B	83 - 86
B-	80 - 82
C+	77 - 79
C	73 - 77
C-	70 - 72
F	< 70

Calendar

The course calendar will be (subject to change at the instructor's discretion): [cs6120_calendar_fall2026](#)

Classroom Recording

This course, or parts of this course, will be recorded for educational purposes. These recordings will be made available only to students enrolled in the course, instructor of record, and any teaching assistants assigned to the course.

If you have objections or would like to opt-out of recordings, please contact the instructor.

Only students who have arranged an accommodation with the Disability Resource Center may use mechanical or electronic transcribing, recording, or communication devices in the classroom. Students with disabilities who believe they may need such an accommodation may contact the Disabilities Resource Center.

Accommodations

It is my job to create a classroom environment that is most conducive to you learning well. If you have accommodations from the [Disability Resource Center](#), please provide your letter to me early in the semester so that I can arrange for these accommodations. If you wish to receive accommodations and do not have a letter, please visit the DRC at 20 Dodge Hall or call (617) 373-2675.

Student Names and Pronouns

We recognize that your legal information doesn't always align with how you identify. Students may update their first and middle names as well as gender marker [with the registrar](#), even if they are not your legal names or gender marker. Those names and gender marker are what would appear publicly in most university systems. In the absence of such updates, what we see on most university systems by default are your legal name and gender marker.

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 does not disrupt teaching or learning. Your comments to others should be constructive and free from harassing statements. You are encouraged to disagree with other students and the instructor, but such disagreements need to be respectful and be based upon facts and documentation (rather than prejudices and personalities). 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.

Part of the learning process in this course is respectful engagement of ideas with others.

The [Code of Student Conduct can be found on the OSCCR website](#).

Title IX

Title IX of the Education Amendments of 1972 protects individuals from sex or gender-based discrimination, including discrimination based on gender-identity, in educational programs and activities that receive federal financial assistance.

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, non-binary, and 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.

Reports can be made 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 "responsible employees" at Northeastern University, meaning they are required to report all allegations of sex or gender-based discrimination to the Title IX Coordinator.

In case of an emergency, please call 911.

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

Religious Holidays

The course staff will make every effort to deal reasonably and fairly with all students who, because of religious obligations, have conflicts with scheduled exams, assignments or required attendance. In this class, contact the course staff at least 7 days in advance of the conflicting date to reschedule a homework or quiz due date.