Applied Natural Language Processing CSE 8803 - Summer 2026

- Course information and success guidance
- Course overview and prerequisites
- Instructor information only (no TA roster)
- Weekly course schedule
- General guidelines, office hours, and inclusion statement
- Grading categories and assessment policies
- Resource collection

Compiled from the course website https://anlp-website.github.io/summer26/

Printable syllabus packet

Edition	Summer 2026
Course	Applied Natural Language Processing (CSE 8803)
Instructors listed on site	Dr. Max Mahdi Roozbahani; Dr. Nimisha Roy; Wafa Louhichi
EdStem / LMS	Canvas dashboard -> Ed Discussion
TA roster handling	TA listings omitted from this packet as requested

This packet compiles the course information, policies, grading structure, schedule, and resource material from the course website into a printable format.

Item	Details
Source website	https://anlp-website.github.io/summer26/
Format	Website-to-PDF syllabus packet
Content basis	Current public course website content, organized into a printable packet
Layout note	Designed for printing and submission while preserving the course website's substantive content

Included sections

Course Information

Instructors: Dr. Max Mahdi Roozbahani; Dr. Nimisha Roy; Wafa Louhichi

EdStem: Check your Canvas Dashboard and click on Ed Discussion.

Course description

The primary objective of this course is to introduce students to broad classes of techniques and tools for analyzing text data using Natural Language Processing (NLP) algorithms and techniques. The course emphasizes how to apply pre-processing, processing, and post-processing NLP techniques to analyze text and develop NLP models.

How to succeed in this class

Engage actively

- Participate in discussions regularly on the class Ed Discussion platform.
- Use office hours to clarify doubts, seek guidance, and deepen understanding.

Review all learning materials

- Review lectures, including videos, tutorial videos, and PDFs.
- Pay close attention to the selected readings because they complement and deepen the lecture content.

Manage assignments strategically

- Start assignments as soon as they are released.
- Use effective time management so the work supports learning, not just completion.

> The site emphasizes that consistent effort, active participation, and timely engagement with course materials are crucial to success.

Course Overview

Course goals

Analyze NLP techniques and apply them to text data. The course is divided into three main categories:

- Preprocessing: demonstrate how to clean and integrate text data.
- Processing: apply NLP algorithms on pre-processed data to perform different tasks.
- Post-processing: evaluate developed NLP models.
- Solve problems with real datasets.
- Gain practical know-how through significant hands-on programming assignments.

Course pre- and/or co-requisites

Students are expected to have prior knowledge in machine learning, linear algebra, optimization, probability, and statistics. The programming language for the class is Python (Python 3.x), and students should know at minimum how to use NumPy, matrix operations, linear algebra, probability, and statistics.

Additional prerequisites

- CSE 6040
- CS 1301

Class text

- Required readings: none.
- Recommended reading: *Introduction to Natural Language Processing* by Jacob Eisenstein (draft available on GitHub).

Instructor Information

Per request, this packet includes the instructor listings from the site and omits the Head TA / TA roster.

Max Mahdi Roozbahani

mahdir@gatech.edu

https://mahdi-roozbahani.github.io/

Wafa Louhichi

wlouhichi3@gatech.edu
https://www.linkedin.com/in/wafa-louhichi/

Nimisha Roy

nroy9@gatech.edu

https://nimisharoy9.wixsite.com/myportfolio

Weekly Course Schedule

The course website states that all deadlines and due dates are in the AoE time zone (23:59 Anywhere on Earth).

WEEK 1 **5/18 - 5/22**	**Topics:** Course introduction; text data preprocessing (normalization, lemmatization, stemming, stop-word removal); text representations: one-hot encoding, bag-of-words, and TF-IDF. **Homework:** HW1 out 5/22. **Quizzes:** Quiz 0 (knowledge-based topics) out 5/18, due 5/22. **Readings:** Chapter 1 and Chapter 2.1 of *Introduction to Natural Language Processing* by Jacob Eisenstein.
WEEK 2 **5/25 - 5/29**	**Topics:** Memorial Day Institute Holiday; classification introduction; Naive Bayes; evaluation with accuracy, precision, recall, and confusion matrix. **Quizzes:** Quiz 1 (week 1 topics) out 5/22, due 5/29. **Readings:** Chapter 2.2 of *Introduction to Natural Language Processing* by Jacob Eisenstein.
WEEK 3 **6/01 - 6/05**	**Topics:** Logistic regression; SVM; perceptron; SVD (dimensionality reduction) and co-occurrence embeddings; GloVe. **Homework:** HW1 due 6/05; HW2 out 6/05. **Quizzes:** Quiz 2 (week 2 topics) out 5/29, due 6/05. **Readings:** GloVe: Global Vectors for Word Representation; Chapters 2.3, 2.4, and 2.5 of *Introduction to Natural Language Processing*.

WEEK 1 **5/18 - 5/22**	**Topics:** Course introduction; text data preprocessing (normalization, lemmatization, stemming, stop-word removal); text representations: one-hot encoding, bag-of-words, and TF-IDF. **Homework:** HW1 out 5/22. **Quizzes:** Quiz 0 (knowledge-based topics) out 5/18, due 5/22. **Readings:** Chapter 1 and Chapter 2.1 of *Introduction to Natural Language Processing* by Jacob Eisenstein.
WEEK 4 **6/08 - 6/12**	**Topics:** Fully connected neural networks; Word2Vec (CBOW and Skip-Gram); toolbox on classification algorithms; toolbox on Word2Vec. **Quizzes:** Quizzes 3 and 4 (weeks 3 and 4 topics) out 6/05, due 6/12. **Readings:** NN Playground; interactive NN initialization; the role of a hidden layer; backpropagation numerical example; more detailed introduction; Efficient Estimation of Word Representations in Vector Space.

WEEK 5 **6/15 - 6/19**	**Topics:** Juneteenth Institute Holiday; CNN; RNN; toolbox on CNN for text classification. **Homework:** HW2 due 6/19, with site note that submission is accepted without penalty until 6/23; HW3 out 6/19. **Quizzes:** Quiz 5 (week 5 topics) out 6/12, due 6/19. **Readings:** CNN live demo; efficient CNN guide and hyperparameter article; backpropagation in CNN; transfer learning in CNN.
WEEK 6 **6/22 - 6/26**	**Topics:** LSTM and GRU; LSTM plus attention with emphasis on the attention mechanism. **Quizzes:** Quiz 6 (week 6 topics) out 6/19, due 6/26.
WEEK 7 **6/29 - 7/03**	**Topics:** Transformer models; examples including BERT and GPT; student recess holiday (7/02); Independence Day Institute Holiday (7/03). **Homework:** HW3 due 7/03, with site note that submission is accepted without penalty until 7/06; HW4 out 7/03. **Quizzes:** Quiz 7 (week 7 topics) out 6/26, due 7/03, with site note that submission is accepted without penalty until 7/06. **Readings:** Attention Is All You Need; BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.
WEEK 8 **7/06 - 7/10**	**Topics:** Sequence labeling: POS tagging; sequence labeling: NER; unsupervised models; topic modeling including latent semantic indexing and LDA. **Quizzes:** Quiz 8 (week 8 topics) out 7/03, due 7/10.

WEEK 9 **7/13 - 7/17**	**Topics:** Introduction to generative AI; prompt engineering techniques; retrieval-augmented generation (RAG); toolbox on exploring LLMs. **Homework:** HW4 due 7/17; HW5 out 7/17. **Quizzes:** Quiz 9 (week 9 topics) out 7/10, due 7/17.
WEEK 10 **7/20 - 7/24**	**Topics:** Foundations of agentic AI; architectures and reasoning in agentic AI; multi-agent and real-world systems; toolbox on agentic AI. **Quizzes:** Quizzes 10 and 11 (weeks 10 and 11 topics) out 7/21, due 7/24.
WEEK 11 **7/27 - 7/28**	**Topics:** Course and homework wrap-up. **Homework:** HW5 due 7/28, with site note that submission is accepted without penalty until 7/31. **Quizzes:** Quiz 12 (weeks 10 and 11 topics) out 7/24, due 7/28, with site note that submission is accepted without penalty until 7/31.

General Guidelines, Office Hours, and Inclusion

Attendance

- This is a fully online course.
- Students should log in regularly so that work can be completed without falling behind and needing extensive review.

Class deliverables

- All class deliverables are handled through Gradescope, except quizzes, which are on Canvas.
- Deadlines will not be extended under any circumstances.
- Work submitted after the deadline receives zero credit.

EdStem

- EdStem is the main and only place for course discussions and announcements.
- If a question is suitable for the full class, students are asked to post it on Ed first so others can benefit and answers arrive faster.
- Sensitive matters may be addressed through a private post; if a more restricted conversation is needed, students can request a private chat.
- The site distinguishes good Ed questions about concepts, assignment interpretation, project questions, site issues, or suggestions from bad questions asking staff to debug code line-by-line.

Exceptional circumstances

- Requests for exceptions should be made in advance when possible and must be due to incapacitating illness, personal emergencies, or similarly serious events.
- The site directs students to work through the Dean of Students office before contacting the teaching team.

Office hours

- TAs plan to hold office hours starting in week 2, except on Georgia Tech holidays.
- Each session is one hour long and is run by at least one TA.
- Office hours will be held through live Ed Discussion threads, with details announced on Ed.
- Students are always welcome to ask questions on Ed; office hours supplement Ed and do not replace it.

Diversity and inclusion

- The course aims to create an environment where all voices are valued and where diversity of gender, sexuality, age, socioeconomic status, ability, ethnicity, race, and culture is respected.
- Students are encouraged to share suggestions that support that goal.
- Conflicts with religious events should be communicated to the instructional team.
- Disability accommodations are handled through Georgia Tech's Office of Disability Services.

Grading Categories and Assessment Policies

Assignments and assessment weights

Component	Weight / Notes
HW1	10% - week 1 to week 2 topics
HW2	15% - week 3 to week 4 topics
HW3	20% - week 5 to week 6 topics
HW4	20% - week 7 to week 11 topics
HW5	15% - week 12 to week 13 topics
Proctored Coding Sessions	5% - three short Honorlock / Vocareum sessions tied to the homework assignments
Quizzes	15% - 12 graded quizzes, lowest grade dropped, all quizzes mandatory

Proctored coding sessions

- Three short proctored coding sessions use Vocareum and Honorlock.
- Each session lasts up to one hour and is designed to take less than the full hour.
- They are intended to assess application of course understanding without relying on generative AI or online lookup tools.

Quizzes

- There are 12 graded quizzes on Canvas, including the knowledge-based quiz (Quiz 0).
- The lowest quiz grade is dropped so that 11 quizzes count toward the grade.
- Each quiz is worth 5 points, and missing a quiz results in a grade of -5 out of 5 rather than 0 out of 5.
- Quizzes are seven minutes long, have five multiple-choice questions, and are mostly conceptual.
- Quiz questions are drawn randomly from a question bank.
- Honorlock is used for quizzes; quizzes are open book and open notes, but additional electronic devices are not permitted and browser activity is restricted to Canvas.

Grading

- Students must achieve an overall weighted average of 60% to pass the course.
- All deliverables are graded by TAs / Gradescope.
- Standard grade thresholds (90, 80, etc.) may be lowered, but not raised.

- Some assignments may include bonus points.

Plagiarism, collaboration, and honor code

- All participants are expected to know and abide by the Georgia Tech Academic Honor Code.
- Plagiarism is a serious offense; submitted work must be the student's own.
- Students may discuss high-level ideas with peers, but each student must write and submit their own answers.
- Students must not place course code in public repositories because future students could copy it.
- Suspected dishonesty is handled through Georgia Tech's academic integrity procedures and can result in severe consequences.

Late policy and due dates

- All homework and quiz deliverables are due at the times shown in the course schedule.
- The course offers no late policy for homework and quizzes.
- Any deliverable submitted after the deadline receives zero credit.
- Students should submit early and verify that every required file is included.

Timing, attendance, and netiquette

- Course videos follow a logical sequence supporting both knowledge-building and experience-building.
- Students are expected to complete assignments by their due dates.
- This is a fully online course, and students should log in regularly.
- The site asks students to follow online discussion ground rules and to communicate respectfully, with correct spelling, punctuation, and grammar consistent with the academic environment.

Resource Collection

Recommended reading

- All content and course materials can be accessed online. The site states that there is no required textbook.
- Georgia Tech students have free access to O'Reilly resources using their official GT email address.

Software engineering / becoming a better programmer

- Debugging
- Clean Code
- Refactoring
- Design Patterns: Elements of Reusable Object-Oriented Software
- The Pragmatic Programmer: From Journeyman to Master

Python

- Python Bootcamp for campus MS Analytics students (by Chris Simpkins).

Data science, machine learning, data mining

- Data Science for Business by Foster Provost and Tom Fawcett.
- The Elements of Statistical Learning: Data Mining, Inference, and Prediction by Trevor Hastie, Robert Tibshirani, and Jerome Friedman.

Probability

- Free probability book by Prof. Guy Lebanon.

Human computation

- Human Computation by Edith Law and Luis von Ahn.

Managing multiple Python versions

- Recommended article: Which Python package manager should you use?
- Managing Multiple Python Versions With pyenv.
- The right and wrong way to set Python 3 as default on a Mac.
- Poetry is highlighted as an increasingly common alternative to conda environments and even setuptools for package workflows.