

LING 582

Advanced Statistical Natural Language Processing

Gus Hahn-Powell

Last Revised October 15, 2026

1 Course Information

This course focuses on statistical approaches to pattern classification and applications of natural language processing to real-world problems.

Course Requisites: LING 539.

Course Objectives

In this course, we will ...

- cover fundamental concepts related to neural approaches to statistical natural language processing, such as ...
 - multi-layer perceptrons (MLPs)
 - convolutional neural networks (CNNs)
 - recurrent neural networks (RNNs)
 - contextual embeddings
 - attention mechanisms
 - transformers
 - large language models (LLMs)
- foster technical skills, such as ...
 - familiarity using the University of Arizona's HPC
 - familiarity using the PyTorch library
 - linux command line basics
 - virtualization and containerization technologies
 - version control (ex. `git`) and the feature branch workflow

Expected Learning Outcomes

Successful students will be able to ...

- write programs ≥ 50 lines of code¹
- read and summarize technical papers in the field²
- use popular NLP libraries for deep learning¹
- use the University of Arizona's HPC to run code¹
- train and use neural networks for NLP tasks³

Credits: 3 units

¹Relates to Linguistics Department's HLT Program Outcome 1.

²Relates to Linguistics Department's HLT Program Outcome 2.

³Relates to Linguistics Department's HLT Program Outcomes 1, 2, & 3.

Prerequisites

- **Successful completion of LING 539**

If you do not meet the prerequisites stated above, you will need the instructor's permission before taking this class.

Instructor

Gus Hahn-Powell

Email: hahnpowell@arizona.edu

Office: Communication Building, Room 316A

Office hours and appointments:

- <https://parsertongue.org/availability/>

Contact

Students should ask all course-related questions in the course forum linked from Class Maestro, where you will also find announcements. For emergencies, or personal matters that you don't wish to put in a private post, please email your instructor at hahnpowell@arizona.edu.

For planning purposes, please note that **your instructor responds to emails and posted questions Monday & Friday between 9AM and 5PM MST**. Typically, you can expect a response within a day.

Schedule

This course consists of 7 Units each lasting approximately a week:

Unit	Topic
1	Getting started/review
2	Introduction to DL for NLP
3	CNNs
4	Tips & tricks for training DNNs
5	RNNs and encoder-decoders
6	Transformers & attention mechanisms
7	Large Language Models (LLMs)

Assessments Please check Class Maestro for assessment details and due dates:

- <https://maestro.hlt.parsertongue.org>

Readings Please check Class Maestro for the readings schedule:

- <https://maestro.hlt.parsertongue.org>

Technology

I strongly recommend that you use a computer with at least 8GB of RAM and ≥ 50 GB of free space.

At the beginning of this course, students will confirm or refresh the development environment used across the HLT sequence. We will use Class Maestro as the course hub, Etude for focused practice and review, Forgejo for course repositories, and the course forum for announcements and questions:

- Class Maestro: <https://maestro.hlt.parsertongue.org>
- Etude: <https://etude.hlt.parsertongue.org>
- Forgejo: <https://git.hlt.parsertongue.org>
- Course forum: <https://forum.hlt.arizona.edu>

Rather than attempt to train neural networks on our local machines, we'll walk through the steps involved in using the <https://public.confluence.arizona.edu/display/UAHPC>.

Python-specific Assistance

For students seeking assistance with Python or data analysis, the University of Arizona Libraries' Data Cooperative provides support through Sarah Stueve, Data Science Specialist. Current support options include Data Drop-in sessions, workshops, and one-on-one consultations:

- Email: sarahstueve@arizona.edu
- Data Cooperative: <https://lib.arizona.edu/research/data>

For students entirely new to Python (but not to programming), this free self-paced course can help bring you up to speed quickly:

- <https://etude.hlt.parsertongue.org>

Other Resources

ResBaz Arizona hosts a variety of events (both virtual and in-person) providing researchers opportunities to connect with experienced data scientists and engineers. For more information, see their homepage:

- <https://researchbazaar.arizona.edu/#portfolio>

Readings

Required readings (papers, chapters, etc.) will be provided by the instructor.

Supplemental Reading

In addition to the course textbook and any posted readings, students may find the following resources useful:

- Dan Jurafsky and James Martin (2024). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. 3rd ed. Upper Saddle River, N.J: Pearson Prentice Hall. URL: <https://web.stanford.edu/%7Ejurafsky/slp3/>
- Jay Alamar and Maarten Grootendorst (2024). *Hands-On Large Language Models*. Sebastopol, California: O'Reilly Media, Inc. ISBN: 9781098150969. URL: <https://github.com/handsOnLLM/Hands-On-Large-Language-Models>
- Yoav Goldberg (2017). *Neural Network Methods for Natural Language Processing*. San Rafael, California: Morgan & Claypool Publishers. ISBN: 1627052984. URL: <https://ebookcentral.proquest.com/lib/uaz/reader.action?docID=4843762>
- Michael A. Nielsen (2015). *Neural Networks and Deep Learning*. Determination Press. URL: <http://neuralnetworksanddeeplearning.com>
- Andrew W Trask (May 2019). *Grokking deep learning*. New York, NY: Manning Publications. URL: <https://www.manning.com/books/grokking-deep-learning>

Code examples will generally be provided in PyTorch (<https://pytorch.org>).

PyTorch Aside from the official tutorials⁴, Students may find the following resources useful:

- Stanford University (2019). *CS230 Introduction to Pytorch Code Examples*. URL: <http://cs230.stanford.edu/blog/pytorch>
- Delip Rao (2019). *Natural Language Processing with PyTorch: Build Intelligent Language Applications Using Deep Learning*. Sebastopol, CA: O'Reilly Media. ISBN: 1491978236. URL: <https://learning.oreilly.com/library/view/natural-language-processing/9781491978221>

Electronic versions of all the recommended resources listed above are freely available to University of Arizona students.

⁴<https://pytorch.org/tutorials/>

2 Evaluation

Your course grade will be based on the following:

Assessment		Description
30%	Course Project	Completed in groups of 1–3 students. The project asks you to define a question or application, select appropriate data and methods, evaluate your approach, and explain what the results do and do not show. Components include reproducible code, a written report, and structured peer feedback on project materials. See Class Maestro for the current instructions, checkpoints, and rubric.
30%	Class Competition	Class-wide text classification competition hosted on Kaggle. Components include a) a written explanation of your approach submitted through Class Maestro, b) reproducible code and setup instructions in the assigned course repository, c) submissions to the class competition leaderboard, and d) participation in peer review of class competition writeups. See Class Maestro for the current invitation, allowed resources, deadlines, and rubric.
25%	Paper Review	Select a technical paper related to NLP, explain its contribution, evaluate the strength of its evidence, and connect it to course concepts. This assignment requires documented use of an approved generative AI system in at least one stage so that you can compare unaided and AI-assisted reading or review workflows. See Class Maestro for the current instructions, checkpoints, and rubric.
10%	Review & Mastery Assignments	Low-risk Etude activities designed to assist in <i>retaining and mastering</i> material covered in each unit.
5%	Misc	Short tasks (check-in survey, repository creation, etc).

Assignment	When?	Where?
Course Project	Part 1: halfway through the term. Part 2: near the end of the term.	See Class Maestro for details.
Class Competition	Part 1: halfway through the term. Part 2: near the end of the term.	See Class Maestro and Kaggle for details.
Paper Review	Part 1: halfway through the term. Part 2: near the end of the term.	See Class Maestro for details.
Review & Mastery Assignments	1-2 weeks after release	Etude, linked from Class Maestro
Misc	Primarily at the start of the course	See Class Maestro for details.

*Dates above are only estimates and are thus subject to change.

Class Maestro is the source of truth for due dates. To help avoid missing important deadlines, I recommend that you subscribe to the course's ICS calendar URL from Class Maestro in the calendar system you are most comfortable using.

3 Grading

Grades and feedback will be posted in Class Maestro:

- <https://maestro.hlt.parsertongue.org>

For assignments involving code, You will be provided with a subset of test cases to help you refine your solution before submitting.

The grading scheme is as follows:

Grade	Point Range
A	90 – 100
B	80 – 89
C	70 – 79
D	60 – 69
E	0 – 59

Grade Disputes

Disputes about grades on a particular project will be entertained for two weeks from the day the project is due, or 1 day before grades are due, whichever is sooner. These will be

resolved by re-grading the entire project. Note that this can result in a lower grade in the event that new mistakes are discovered.

No negotiations about individual students' letter grades will be entertained once final grades are assigned, except as permitted by the policy stated above.

Collaboration Policy

Students are encouraged to discuss problems, general approaches, sources, tools, and debugging strategies. Project implementations and the associated documentation for each project must be completed by the individual or team submitting that work. **Copying another person's work, submitting work from another team as your own, or reusing uncited material from a website, repository, paper, or generative AI system is not permitted and will be treated as a case of academic dishonesty.**

Use of Generative AI

This course studies modern NLP systems, including large language models, so generative AI tools are part of the subject matter as well as part of the contemporary technical workflow. Some assignments will require generative AI use, such as AI-assisted paper review, tool-calling activities, structured-output practice, or model-system evaluation. When an assignment requires AI use, the goal is to make the system's behavior inspectable: what you asked, what it returned, what you verified, what failed, and how that changed your own analysis.

Unless explicit permission is otherwise given, students may only use open source, open weight models in this course. Within that constraint, and unless a specific assignment says otherwise, students may use generative AI tools for brainstorming, planning, coding support, debugging, revision, and explanation. Any substantive use must be disclosed in the submitted work, including the tool or model used and enough detail about the interaction for the instructor to understand how it shaped the submission. For work involving model outputs, tools, retrieval, structured output, or evaluation, assignment instructions may require prompts, model/version information, tool definitions, output examples, validation results, or a brief account of failed attempts.

Students remain responsible for the correctness, originality, security, and reproducibility of their work. Generated text, code, citations, analyses, labels, or leaderboard strategies should be inspected rather than treated as authoritative. Do not submit private, confidential, credential-bearing, or personally sensitive material to a model unless the assignment explicitly authorizes the data and tool. Assignment-specific limits take precedence over this general policy.

Late Policy

Projects are due electronically via the system named in Class Maestro by the stated deadline. No late work will be accepted.

4 University Policies

Missed Classes (Absence)

Accommodation of Religious Observance and Practice:

- <https://policy.arizona.edu/human-resources/religious-accommodation-policy>

All holidays or special events observed by organized religions will be honored for those students who show affiliation with such religions. Absences pre-approved by the UA Dean of Students office will be honored. No matter the reason for missing class, the student is always responsible for the missed material.

With the exception of the above, attendance⁵ is mandatory. Students who miss more than 1/3 of classes will be dropped.

Classroom Behavior

To foster a positive learning environment, students and instructors have a shared responsibility. We want a safe, welcoming, and inclusive environment where all of us feel comfortable with each other and where we can challenge ourselves to succeed.

The Arizona Board of Regents Student Code of Conduct is here:

- <https://public.azregents.edu/Policy%20Manual/5-308-Student%20Code%20of%20Conduct.pdf>

ABOR Policy 5-308, prohibits threats of physical harm to any member of the University community, including to oneself:

- <http://policy.arizona.edu/education-and-student-affairs/threatening-behavior-students>

Valuing Inclusion and Diversity

The University of Arizona strives to foster inclusive learning environments in which diversity is recognized and valued. This course strives to create a civil and welcoming environment for everyone, including students of diverse ethnic, cultural, linguistic, national, and familial backgrounds and gender identities, ages, abilities, and veteran status. Students and faculty are responsible for creating an inclusive learning environment through respectful and civil discussion in the classroom and inclusive practices in class and group work.

- We will use the names and pronouns as selected and proposed by each individual in the course in acknowledgment of the intricate nature of our identities and in respect of each other's integrity. Students may request a University wide name change through LGBTQ Affairs at and learn more about name and pronoun use at <https://lgbtq.arizona.edu/transgender-resources>.

⁵For students in online-only classrooms, attendance means timely completion of assignments and modules as outlined in the course schedule.

- If you have questions about particular policies concerning gender equity, sexual harassment, or sexual assault, please contact the Office of Institutional Equity (<https://titleix.arizona.edu>) or the Dean of Students Office (<https://deanofstudents.arizona.edu>).

Respectful and Careful Communication

All communication in this class adheres to the principles of civil discourse. Civil discourse is guided by mutual respect and appreciation. Diversity of knowledge is an asset to class discussions. In all communication, you are expected to be scholarly, professional, and respectful. Constructive criticism in discussion of course concepts is highly encouraged. Mocking and/or bullying are never allowed. To be critical does not exclude being polite. See:

- UA National Institute of Civil Discourse: <https://nicd.arizona.edu>
- UA Dean of Student's page on cyberbullying: <https://deanofstudents.arizona.edu/safety/cyberbullying>

Accessibility and Accommodations

At the University of Arizona, we strive to make learning experiences as accessible as possible. If you anticipate or experience barriers based on disability or pregnancy, please contact the Disability Resource Center (520-621-3268, <https://drc.arizona.edu>) to establish reasonable accommodations.

Student Code of Academic Integrity

Students are encouraged to share intellectual views and discuss freely the principles and applications of course materials. However, graded work must be the product of independent effort unless otherwise instructed. **In this course, “independent” means that submitted work must honestly represent the contribution of the submitting individual or team and must disclose substantive assistance from people, tools, sources, or generative AI systems as required by this syllabus and the assignment instructions.** Students are expected to adhere to the UA Code of Academic Integrity as described here:

- <https://deanofstudents.arizona.edu/policies/code-academic-integrity>

Confidentiality of Student Records

<https://registrar.arizona.edu/personal-information/student-information>

On Dropping Classes

If you find yourself thinking about dropping this (or any other) class, first make sure that that's what you really want to do. Chatting with the instructor or your academic advisor may help. If you drop within the first week of the term, there will be no notation on your

transcript; it will be as though you'd never enrolled. After the second week, a drop will be recorded on your transcript. You will receive a "WP" (withdrawn passing) only if you were passing the class at the time of your drop. Toward the end of the term, dropping becomes a challenge, because you need to explain to the instructor and to the dean why you were unable to drop the class during the first half of the term. For drop deadlines specific to this compressed format, please see this calendar:

- <https://www.registrar.arizona.edu/dates-and-deadlines>

Subject to Change Statement

The instructor reserves the right to change with advance notice where appropriate the content of the course. This right does not apply to posted grading and absence policies or University Policies.