





# CSCI 4140: Natural Language Processing

## CSCI/DASC 6040: Computational Analysis of Natural Languages

Spring 2024: Jan 8 – May 6  
3 Credits

Contact me 	Meet with me 	Class location and time 	Course materials 
Nic Herndon <a href="mailto:herndonn19@ecu.edu">herndonn19@ecu.edu</a>	Tue 9:00 – 10:00 & 4:00 – 5:00 Thu 9:00 – 10:30 & 3:30 – 5:00 SciTech C-108 and <a href="#">MS Teams</a>	TTh 2:00 – 3:15 PM <a href="#">Bate 1028</a> and <a href="#">MS Teams</a>	Listed in Section <a href="#">1.3</a>

## Contents

<b>1 Course Description</b>	<b>1</b>
1.1 Course Objective	2
1.2 Prerequisites	2
1.3 Course materials	2
1.4 Tentative Schedule	3
<b>2 Course Staff</b>	<b>3</b>
<b>3 How to succeed in this course</b>	<b>4</b>
3.1 Feedback	4
3.2 Equality Statement	4
3.3 Accommodations	4
3.4 Student Hours	5
3.5 Attendance and Participation	5
3.6 Missed Instructional Time in the Event of a Disruption	5
3.7 Course Staff Responsibilities	5
<b>4 Grading and Course Assignments</b>	<b>6</b>
4.1 Regrade Requests	6
<b>5 Final Grades</b>	<b>6</b>
5.1 FAQ regarding final grades	6
<b>6 Course Incompletes</b>	<b>7</b>
<b>7 Academic Honesty</b>	<b>7</b>
<b>8 Class Recordings</b>	<b>7</b>
<b>9 Final Note</b>	<b>7</b>

## 1 Course Description

Natural Language Processing (NLP) is the engineering art and science of how to teach computers to understand human language. NLP is a type of artificial intelligence technology, and it's now ubiquitous – NLP lets us talk to our phones, use the web to answer questions, map out discussions in books and social media, and even translate between human languages. Since language is rich, ambiguous, and very difficult for computers to understand, these systems can sometimes seem like magic – but these are engineering problems we can tackle with data, math, and insights from linguistics.

## 1.1 Course Objective

This course will introduce NLP methods and applications including probabilistic language models, machine translation, and parsing algorithms for syntax and the deeper meaning of text. During the course, students will (1) learn and derive mathematical models and algorithms for NLP; (2) become familiar with key facts about human language that motivate them, and help practitioners know what problems are possible to solve; and (3) complete a series of hands-on projects to implement, experiment with, and improve NLP models, gaining practical skills for natural language systems engineering.

By the end of the semester students will:

- Acquire the fundamental linguistic concepts that are relevant to language technology.
- Analyze and understand state-of-the-art algorithms and statistical techniques for reasoning about linguistic data.
- Implement state-of-the-art algorithms and statistical techniques for reasoning about linguistic data.
- Adapt and apply state-of-the-art language technology to new problems and settings.
- Read and understand current research on natural language processing.
- Exercise public speaking skills by discussing various course topics with classmates, asking and answering questions, and doing group presentations.

By the end of this course students should be able to transfer the knowledge gained, and apply it outside of the context of the course to:

- Identify, formulate, analyze, and solve complex computing or engineering problems by applying principles of computing, engineering, science, and mathematics.
- Design, implement, and evaluate a computing or engineering solution to meet a given set of requirements, with consideration of public health, safety, and welfare, as well as global, cultural, social, environmental, and economic factors.
- Communicate effectively in a variety of professional contexts, with a range of audiences.
- Function effectively as a member or leader of a team engaged in activities appropriate to the program's discipline, creating a collaborative and inclusive environment, establishing goals, planning tasks, and meeting objectives.

## 1.2 Prerequisites

- For CSCI 4140: CSCI 2540 Data Abstraction and Object-Oriented Data Structures, MATH 2228 Elementary Statistical Methods I, or MATH 2283 Statistics for Business.
- For CSCI/DASC 6040: enrolled in the master of science in computer science, data science, or software engineering programs or consent of instructor.

## 1.3 Course materials

- Optional: *Speech and Language Processing* (SLP) – an introduction to natural language processing, computational linguistics, and speech recognition.
- Optional: *Dive into Deep Learning* (D2L) – interactive deep learning book with code, math, and discussions.

## 1.4 Tentative Schedule

Date	Topic
Tue 01/09	Introduction
Thu 01/11	N-gram LMs - <b>Homework 1</b> on N-gram LMs released (due Sun 1/21) - Optional: Introduction to Python (videos)
Tue 01/16	Vector embeddings - <b>Homework 2</b> on Exploring Word Vectors released (due Sun 2/04) - <b>Homework 3</b> on word2vec released (due Sun 2/18)
Thu 01/18	Steamship LMMs - <b>Homework 4</b> on LMM applications released (due Sat 3/02)
Tue 01/23	Artificial neural networks
Thu 01/25	<b>Final project details</b> (video; no class)
Tue 01/30	Recurrent neural networks
Thu 02/01	PyTorch tutorial
Tue 02/06	Fancy RNNs (videos; no class) - <b>Homework 5</b> on Neural Machine Translation with RNNs and Analyzing NMT Systems released (due Sun 3/17)
Thu 02/08	Self-attention and transformers (videos; no class)
Tue 02/13	Hugging face tutorial
Thu 02/15	Nano GPT - <b>Homework 6</b> on Nano GPT released (due on Sun 3/31)
Tue 02/20	Pretraining - <b>Homework 7</b> on Self-Attention, Transformers, and Pretraining released (due on Sun 4/14)
Thu 02/22	Guest lecture on public speaking
Tue 02/27	Neural language generation
Thu 02/29	Prompting, instruction fine-tuning, and RLHF
Sun 03/03	Spring break
Sun 03/10	Spring break
Tue 03/12	Qestion answering
Thu 03/14	CNN, TreeRNN
Tue 03/19	Code generation
Thu 03/21	Multimodal deep learning
Tue 03/26	Model analysis and explanation
Thu 03/28	Model interpretability and editing
Tue 04/02	Ethical issues related to NLP
Thu 04/04	Course wrap-up
Tue 04/09	Final project presentations
Thu 04/11	Final project presentations
Tue 04/16	Final project presentations
Thu 04/18	Final project presentations

## 2 Course Staff

My name is Nic Herndon. I worked in industry and academia, in both public and private sectors, and this is the fifth time I teach this course. Assisting me this semester are Madison Rose ([rosem19@students.ecu.edu](mailto:rosem19@students.ecu.edu)) and Cris Zbavitel ([zbavitelc19@students.ecu.edu](mailto:zbavitelc19@students.ecu.edu)).

### 3 How to succeed in this course

Your success in this class is important to me. We all learn differently and bring different strengths and needs to the class. If there are aspects of the course that prevent you from learning or make you feel excluded, please let me know as soon as possible. Together we'll develop strategies to meet both your needs and the requirements of the course. There are also a range of resources on campus, including:

- [Center for Communication Excellence](#) – can help with managing speaking anxiety, working on a professional social media presence, developing professional interpersonal communication skills, and developing engaging public speaking skills. Do you wish you felt more confident when communicating with your professors and supervisors? Do you dread getting up to speak in front of your peers? Would you like to be a more engaging speaker when giving class presentations? Do your professional communication skills need help? The Center for Communication Excellence, located in 205 Joyner East, is available to you as a FREE service. Face to face and virtual appointments are available. For more information and hours of operation, call 252.328.2790 or email [commcenter@ecu.edu](mailto:commcenter@ecu.edu).
- [Computer Science Peer Tutoring](#) – peer tutoring for a number of undergraduate courses in Computer Science and Software Engineering. The tutoring schedule for the current semester can be found [online](#).
- [Pirate Academic Success Center](#) – the center offers free tutoring for science and humanities courses, a variety of math and business courses, foreign languages, and study skills such as time management, test taking, and effective textbook reading. Call (252) 737-3009, email [tutoring@ecu.edu](mailto:tutoring@ecu.edu) or visit 2300 Old Cafeteria Complex to make an appointment. All services are free to ECU students.
- [Counseling Center](#) – provides individual counseling within a short-term therapeutic model, with most students attending 4-5 sessions.
- [ECU CARES](#) – offers assistance to distressed individuals, connecting them to appropriate campus resources, and reporting concerning behavior to professionals on campus. This can be a critical first step in helping the individual(s) improve and ensuring a safer campus for everyone.
- [Language Academy](#) – a premier destination for students interested in improving their English skills and understanding of American culture.

#### 3.1 Feedback

I encourage every student to [submit anonymous feedback](#) throughout the semester on how to improve this course. In addition, during the lectures we'll be using [this etherpad](#) which can be used to post anonymous questions and to chat with classmates.

#### 3.2 Equality Statement

I am dedicated to establishing a learning environment that promotes diversity of the students including race, class, culture, religion, gender, sexual identity, and physical ability. It is important that this is a safe classroom environment. We will practice being generous and respectful members of our class and computer science community. Please let me know immediately if you notice discriminatory behavior in this class or feel discriminated against.

#### 3.3 Accommodations

East Carolina University seeks to comply fully with the Americans with Disabilities Act (ADA). Reasonable accommodations will be made for students with verifiable disabilities. In order to take advantage of available accommodations, students must be registered with the Department for Disability Support Services located in Slay 138, 252-737-1016. Additional DSS student resources can be found at: <https://accessibility.ecu.edu/students/>.

### 3.4 Student Hours



Student hours are an important part in supporting you throughout this course. Even if you don't have specific questions, needs, and concerns, I would love to meet up with you at least once during this semester. There are a couple of options to meet up: meet with me using [MS Teams](#) during the student hours, or, if these times don't work for you, please email me and I will find another day and/or time. You are also welcome to reach out to the teaching assistants with any questions related to this course.

### 3.5 Attendance and Participation

- Face-to-face students are expected to attend, and be on time, for every class. This demonstrates professionalism and consideration for your fellow students and your instructor. While the course does not have an attendance policy, students who miss class and/or are late for class may experience an impact on their grade by missing classroom activities.
- Distance education students are expected to demonstrate professionalism and courtesy by muting their microphones during lectures streamed live, and unmute them when participating in class activities (Q&A, ICEs, etc.).
- For group work (in-class presentation) students are expected to clearly state to their teammates, their expectations on communications (in person, email, Microsoft Teams, etc.), deadlines, and contributions of each student. If the agreed expectations are not met by one of the students, I encourage the teammates to reach out to him/her, and if a resolution cannot be reached, to inform me right-away.
- Students should make every effort to participate in class activities.
- Missing class does not automatically result in extensions on assignments or exam due dates.
- There is a clear correlation between attendance/participation and your grades.
- It is the student's responsibility to seek out the instructor and other classmates to obtain the information (e.g., lecture notes, relevant announcements, etc.) if they missed class.
- Students should maintain regular communication with instructors regarding their health status and communicate any changes in their ability to complete coursework and academic responsibilities immediately.
- You may at any time consult with your advisor or the instructor about whether to request an Incomplete ([ECU Faculty Manual IV.VIII.E Grade of Incomplete](#)) or a Withdrawal ([ECU's Withdrawals Policies](#)).

### 3.6 Missed Instructional Time in the Event of a Disruption

Making up missed instructional time in this course will follow [ECU's Policy for Making Up Missed Instructional Time Due to Suspension of Instruction](#).

### 3.7 Course Staff Responsibilities

You can expect from me:

1. Timely release of course assignments.
2. Timely release of scores achieved on course assignments.
3. To respond to questions in the discussion forums in a reasonable amount of time.
4. Be respectful of your ideas and value the diversity you bring to the class.
5. Be open to dialogue that challenges me.
6. Be present during the stated office hours.
7. Ensure the proper running of the course.

## 4 Grading and Course Assignments

**Homework (70%):** There will be seven homework assignments on topics covered in class. Everyone will get a total of three late calendar days to use for homework assignments. After all three late days have been exhausted, no more late submissions will be accepted. For unforeseen health and personal emergencies, please contact me. Job interviews/other schoolwork are not excuses for late homework. The homework assignments will be posted on the course page in [Canvas](#), and have to be submitted via Canvas. They are due on the dates listed in the tentative schedule by 11:59 PM.

**Final project (30%):** There will be several assignments as part of the final project, including but not limited to, discussions, teamwork, project proposal presentation, and final project presentation.

### 4.1 Regrade Requests

If you feel you deserved a better grade on an assignment, you may submit a regrade request via email within three days after the grades are released. Your request should briefly summarize why you feel the original grade was unfair. Your TA will reevaluate your assignment as soon as possible, and then issue a decision. If you are still not happy, you can ask for your assignment to be regraded by an instructor. If no request was received within three days, the grade remains final for that assignment.

## 5 Final Grades

To evaluate your understanding of the course content I will use scores achieved on each of the above assessment components. Your final grade will convey what you know from the course and how well you know it. Missing and late assignments can have a dramatic impact on your final grade so it is important that you are attentive to submission deadlines and avoid any missing work. The typical breakdown of percentages and final grades for this course are as follows.

- For undergraduate students:

F	D	D+	C−	C	C+	B−	B	B+	A−	A
0 – 59	60 – 66	67 – 69	70 – 72	73 – 76	77 – 79	80 – 82	83 – 86	87 – 89	90 – 92	93 – 100

- For graduate students:

F	C	B	A
0 – 69	70 – 79	80 – 89	90 – 100

This grading scheme may be adjusted based on the overall performance of students in the course.

### 5.1 FAQ regarding final grades

1. Will there be extra credit offered at the end of the semester?  
No. There will be extra credit offered throughout the semester – please take advantage of it.
2. But I had difficulty during the semester due to \_\_\_\_\_. Isn't there something that can be done?  
Please notify me right away if you have any emergencies; don't wait until the end of the semester to bring this to my attention.
3. My final grade is 89.99%. Will I get an A?  
No. The grade ranges are listed above, and there will be no rounding up. 89.99% earns you a B if you're a graduate student, and an B+ if you're an undergraduate student.
4. I turned in my homework late by 1 minute because the Internet was slow. Will that use any of my late days?  
Yes. Late days are used in one day increments – turning in a homework between 1 second and 24 hours late counts as one late day. Each homework assignment has unlimited submission attempts; so, turn it in early, and turn it in often, just as you would check your code in a `git` repository. We will grade your latest homework submission.

5. I used my three late days for homework 1. Can I use three late days for homework 2?  
No. Each student has a total of three late days for the whole semester. Once you used them you have to turn in on time the remaining homework assignments.
6. Is it okay to work alone without forming a team?  
No. In a professional setting you never work alone (you might work by yourself, but you're a part of the team). Learning how to work well in a team setting is a must.

## 6 Course Incompletes

Students who are unable to complete course requirements within the allotted time because of severe medical or personal problems may request a grade of Incomplete from the instructor of the course. Incomplete grades are warranted only if a student is passing the course at the time of the request and if the course requirements can be completed by the end of the following semester.

**Note:** an incomplete means you are on your own to complete the material agreed upon by the instructor of this course. Do not expect additional help or one-on-one teaching of the material past the course completion date. It is your responsibility to complete the remaining material.

## 7 Academic Honesty

It is very important in all courses that you are honest in all the work that you complete. You may discuss assignments with other students, in fact I encourage this as a learning experience. But again, the writeup must be your work. Copying is not allowed, and collaboration so close that it looks like copying, is not allowed. Remember to tell me who you worked with as well.

If you copy on project assignments or exams you are doing a disservice to yourself, the instructor for the course, the Department of Computer Science, the East Carolina University, and your future. We design our courses to provide you the necessary understanding and skill that will make you an excellent computer scientist. Assignments are designed to apply and test your knowledge and understanding of the material.

I will carefully review your submissions automatically and manually to verify that “cheating” has not taken place. If you are suspected of plagiarism, I will follow an informal path to determine if academic dishonesty has taken place, and you may receive an F for the course and have a mark on your permanent record at ECU. This will disrupt your schedule for completing courses and may lead to you not completing your degree in a timely fashion. Please review carefully the [Academic Integrity](#) to understand what academic dishonesty is, how you can avoid it, and the procedure I will follow if you are under suspicion. If you have questions or unsure if something constitutes plagiarism, please reach out to me.

Every assignment implicitly includes a “contract” that you sign virtually by submitting your assignment. By “signing” this contract you indicate that you have read all the documents on this website and any links to academic honesty associated with the university. Your signature indicates that you completely understand the policies in place and that you have not plagiarized.

## 8 Class Recordings

This class will be recorded and broadcast on the internet and/or distributed on other electronic media now or hereafter known. These recordings may contain your image and your voice. You must notify me as soon as possible if you DO NOT want your image and your voice contained on the recording.

## 9 Final Note

Occasionally, it may be necessary to revise this syllabus due to extenuating circumstances. I reserve the right to revise this syllabus if the need arises. If I do so, I will provide you with advance notice.