

**Natural Language Processing
CAI 6307**

Section: 1XYZ, OVER [FTF and Virtual] (30041, 30592)

Class Periods: Mondays Periods 8-10 (3:00 PM – 6:00 PM)

Location: [MCCB G086](#) and Zoom [see weekly Canvas Modules]

Academic Term: Spring, 2024

Instructor:

Sangpil Youm, youms@ufl.edu (with Prof. Bonnie J Dorr)

Office Hours: By appointment (request 24 hours in advance), zoom link distributed within Canvas

Teaching Assistants and Graders/Peer Mentors:

Chathuri Jayaweera (lead TA), chathuri.jayawee@ufl.edu

Venkata Pamarthy, v.pamarthy@ufl.edu

Sudesh Pawar, sudeshpawar@ufl.edu

Divyansh Singh, divyansh.singh@ufl.edu

Uzair Fasih, mfasih@ufl.edu

Justin Ho, justinho@ufl.edu

Eugene Li, lieugene@ufl.edu

Course Description

Covers concepts in natural language processing ranging from shallow bag-of-words to richer representations and formalisms, for applications such as translation, generation, extraction, summarization, and dialogue. Classic and state-of-the-art techniques and remaining challenges are discussed, as well as recent proposals for meeting those challenges (symbolic, machine learning, and generative models). Intended for graduate students doing research related to natural language processing.

Course Pre-Requisites / Co-Requisites

Prerequisites: Proficiency in programming (Python recommended) & familiarity with introductory machine learning or artificial intelligence is a plus.

Course Objectives

Students will:

- Learn advanced concepts and their application in natural language processing (NLP) through **weekly readings and assessments**, in preparation for a course project.
- Develop foundational knowledge of the most recent technologies and advances, including neural, attention-based, and transformer models, through instructor-produced **videos and quizzes** before each class period, coupled with in-class **discussion**.
- Gain **proficiency in multiple topical areas/applications**, e.g., text processing and n-grams, linguistic structure and analysis, semantics-driven applications, sequence labeling applications, sequence-to-sequence mappings, and techniques for evaluating NLP output.
- Learn how one **designs, critically assesses, and systematically evaluates** solutions for different NLP applications.
- Engage in a **team-driven project** chosen from a range of topical areas/applications, including:
 - Oral delivery of a **project proposal presentation**.
 - Oral delivery of a **final project presentation**.
 - Completion of a **final project paper** with literature review, algorithm design, experimentation, and assessment using sound metrics.

Materials and Supply Fees There are no materials or supply fees for this course.

Required Textbooks and Software

All readings are online and freely available:

- Title: *Speech and Language Processing* (third edition, advanced chapters)
- Author: Daniel Jurafsky and James Martin
- Date: 2023 (updated regularly at <https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf>)
- Textbook is referred to as “SLP” below.

Supplemental Readings and Resources:

Recommended online, selected readings (freely available):

- Proceedings of ACL, EMNLP, NAACL, for example: <https://aclanthology.org/events/acl-2023/>
- Association for the Advancement of Artificial Intelligence (AAAI): <http://www.aaai.org/AITopics/pmwiki/pmwiki.php/AITopics/NaturalLanguage>
- The second edition of *Natural Language Processing with Python* is published here: <http://www.nltk.org/book/>

Course Schedule

Here is a general schedule for the semester. Please refer to the *Canvas Announcements* and *Modules* for final details about each topic, exact dates, and the specific schedule we implement this semester.

Themes	Weeks	Topics	Readings & Assessments
Introduction	Week 1 1/8	Introduction, Motivation, Application examples, LLMs	Slides/Videos – Wk0, Wk1 [Asst 1 out]
Topic 1: Text Processing, Corpora, Words, and N-grams & Pitches			
Pipeline and Preprocessing; Research Project Planning [MLK: no class on 1/15]	Week 2 1/16	NLP Pipeline and pre-processing	Slides/Videos – Wk2 SLP 2.1-2.2
		NLP Applications and Research Project Planning	Proceedings of ACL, EMNLP, NAACL [Asst 1 due Wed 11:59pm] [Asst 2 out]
Words and Corpora; Pitches/Speeches	Week 3 1/22	Text Processing in Python	Slides/Videos – Wk3 SLP 2.3, 2.4 [Quiz 1 due Mon 3pm] [Quiz 2 due Mon 3pm]
		Project/Pitch Planning	[Initial Team sign-up in Canvas: Mon, 6pm] [Asst 2 due Wed 11:59pm] [Asst 3 out]
Language Modeling; Pitches/Speeches (cont)	Week 4 1/29	N-gram Language Models	Slides/Videos – Wk4 SLP 3.1-3.5 [Quiz 3 due Mon 3pm] [Continue Team sign-up]
		<i>Project Pitches/Speeches #1</i>	
Topic 2: Linguistic Structure and Analysis & Proposal Prep			
Grammars and Parsing; Prep for Project Proposals	Week 5 2/5	Constituency Grammars, Parsing	Slides/Videos – Wk5 SLP 17.1-2, 17.5 [Quiz 4 due Mon 3pm]

		Project Pitches/Speeches #2	[Finalized Teams Fri 11:59pm] [Asst 3 Wed 11:59pm] [Asst 4 out]
Ambiguity, CKY, Dependency Parsing	Week 6 2/12	Ambiguity, CKY Dependency Tree Parsing	Slides/Videos – Wk6 SLP 17.4, 17.6, 18.1-18.2 [Quiz 5 due Mon 3pm] [Asst 4 due Wed 11:59pm] [Proposal Asst out]
Topic 3: Sequence Labeling & Proposal Presentations			
Sequence Labeling; NLP Proposals I-IV	Week 7 2/19	Sequence Labeling for POS and NE	Slides/Videos – Wk7 SLP 8.1-8.3, 17.3 [Quiz 6 due Mon 3pm]
		Proposal Talks A	[Proposal Asst A: Slides due Fri 2/16 @11:59pm, Red-team Q due Mon 2/19 noon]
	Week 8 2/26	HMMs, Viterbi, Conditional Random Fields	Slides/Videos – Wk8 SLP 8.4-8.5 [Quiz 7 due Mon 3pm]
		Proposal Talks B	[Proposal Asst B: Slides due Fri 2/23 @11:59pm, Red-team Q due Mon 2/26 noon] [Asst 5 out]
Topic 4: Semantics I - Vector representations & Embeddings (& Metrics Intro)			
Semantics I; Prep for Final Talks <i>[Spring Break: No class 3/11-3/15]</i>	Week 9 3/4	Word Meaning, Vector Semantics and Embeddings	Slides/Videos – Wk9 SLP 6.1-6.3 [Quiz 8 due Mon 3pm]
	Week 10 3/18	Word Similarity; Cosine, W2V, TFIDF, Jaccard, Intro to Metrics	Slides/Videos – Wk10 SLP 6.4, 6.5, 6.8, 4.7 [Quiz 9 due Mon 3pm]
		Prep for Project presentations	[Asst 5 due Wed 11:59pm] [Presentation Asst out]
Topic 5: Semantics II - Semantic Analysis and Role Labeling & Presentations			
Semantics II; Final Talks I-V	Week 11 3/25	Semantic Role Labeling	Slides/Videos – Wk11 SLP 24 [Quiz 10 due Mon 3pm]
	Week 12 4/1	Contextualized Embeddings	Slides/Videos – Wk12 SLP 11.1-11.3 [Quiz 11 due Mon 3pm] [Presentation Asst A: Slides due Fri 3/29 @11:59pm, Red-team Q due Mon 4/1 noon]
Project Presentations A			

Topic 6: Sequence-to-Sequence Mappings & Presentations			
Sequence-to-Sequence; Project Papers Underway	Week 13 4/8	Machine Translation, Encoder-Decoders, Transformers	Slides/Videos – Wk13 SLP 13.1, 13.5, 10.2, 10.4 [Quiz 12 due Mon 3pm] [Presentation Asst B: Slides due Fri 4/5 @11:59pm, Red-team Q due Mon 4/8 noon] [Checkpoint #1 Asst out]
		Project Presentations B	
Topic 7: Evaluation and Metrics & Project Papers underway			
Evaluation and Metrics	Week 14 4/15	Evaluation Metrics, Guest Lectures TBD	[Work on Project papers] [Checkpoint #1 due Mon 3pm; Checkpoint #2 Asst out]
Final Papers			
Final Meeting	Week 15 4/22	Guest Lectures TBD & Final paper discussions	[Checkpoint #2 Due Mon 3pm; Final Papers Due 4/24 11:59pm]

Attendance Policy, Class Expectations, and Makeup Policy

Class time occurs *synchronously* in our assigned face-to-face classroom [F2F] and virtually [online] via **Zoom: Mondays Periods 8-10 (3:00 PM – 6:00 PM EST)**.

Format: This course is taught as a “flipped classroom.” All lectures are pre-recorded and delivered for viewing outside of the classroom in the form of brief (10-20 min) video snippets. These videos are accompanied by required “quizzes” to solidify understanding of the lecture material. You are responsible for watching all videos (and taking all quizzes) relevant to the next class the start of class each week.

Class time (periods 8-10) will also be recorded and is dedicated to: (a) exploration of advanced concepts presented in the pre-recorded lecture, (b) review of assigned material and module-relevant discussion; (c) identification of problems in a variety of application areas to be tackled in students’ course projects; (d) presentation (and critique) of project proposals; (e) presentation (and critique) of project hypotheses, findings, outcomes, limitations, and conclusions. In order to be successful in this course, you are advised to attend class sessions (either F2F or virtually) and to take advantage of this unique flipped classroom opportunity.

General Attendance:

- In-class attendance is not required, as there will be a virtual option by zoom.
- However, it is best to appear in person for presentation of project proposals and final presentations. You will be expected to critique presentations by fellow classmates.
- You are responsible for all material covered during pre-recorded lectures and class discussions.
- All class (FTF) periods will be recorded (see the disclaimers and notice that follow).

Disclaimer #1: Many Zoom sessions are being held campus-wide to deliver courses and the demand for recorded materials is high. It is possible there will be delays while completing the recording process. *I will post live classroom recordings as soon as possible [under the corresponding Canvas Module],* however the way to ensure that you see the classroom activities is to participate in the

synchronous FTF classroom meeting on Tuesdays.

Disclaimer #2: The Zoom platform records these sessions and I cannot guarantee they will not experience errors in the recording process. Please be aware that attending the live classroom sessions is always the best way to ensure you are able to view the complete content of the discussion.

Notice: Our classroom discussion is audio visually recorded for students in the course to reference after the live recorded session. Students who choose to participate with their camera engaged or utilize a profile image are agreeing to have their video or image recorded. If you are unwilling to consent to have your profile or video image recorded, be sure to keep your camera off and do not use a profile image. Likewise, students who un-mute during class and participate orally are agreeing to have their voices recorded. If you are not willing to consent to have your voice recorded during class, you will need to keep your mute button activated. A textual “chat” feature is also available during FTF class time. “Chats” sent through the public channel may be recorded. If you do not wish for your “chat” to be recorded, you must send it through the private channel directly and only to the instructor. Please note, since the FTF class time is a public setting, private chats may or may not always be answerable during the live class time.

Excused absences: Excused absences must be consistent with university policies in the Graduate Catalog (<https://catalog.ufl.edu/graduate/regulations>) and require appropriate documentation. Additional information can be found here: <https://gradcatalog.ufl.edu/graduate/regulations/>

Makeups or extensions: Makeups or extensions for graded class activities are provided given appropriate documentation is presented in a timely fashion; however, keep in mind that flexibility is already built into the flipped classroom schedule, to avoid the need for makeups or extensions. For example, quizzes and assignments may be done in a time period that you select within a multi-day window. Moreover, the lowest four quiz scores are dropped. This approach allows you to plan ahead and work on assessments within a timeframe that easily accommodates scheduling constraints.

Late Submissions: Quizzes are due at 3pm on the day of class. No quiz can be accepted after that time as the solution is discussed during the 3-6pm class period on that same day. For assignments: If the solution has not been discussed and you submit one day late, there will be a penalty of 50% on the graded work. For example, if an assignment is due on a Wednesday at 11:59pm, then: (a) At 12:00 midnight (one minute later) the assignment is one day late; (b) Work that is two or more days late (any time after 11:59pm Thursday night) will **not** be accepted.

Questions outside of class time: It is possible to set up a zoom meeting (via Canvas) with the instructor or TA for private discussions that cannot be handled during a public classroom session. This may be needed, for example, to finalize the details of a team project proposal. For questions related to concepts presented in FTF/virtual meetings and/or pre-recorded lectures, at least one hour is set aside each week during FTF class time (in a group setting) in which you may raise discussion points or questions, so that these are discussed in a setting where your classmates also benefit from that discussion. This also reduces redundancy, as many students have the same questions of interest to the entire class. Please consider taking advantage of this unique flipped classroom experience before contacting the instructor or TA to meet outside of the classroom for questions regarding assessed concepts (quizzes and assignments) that are common to the entire class.

Evaluation of Grades

Assessments are scheduled over multi-day time windows: quizzes, assignments, and project (proposal, presentation, paper). The “flipped classroom” format provides ample time for the entire class to prepare interactively (together) for assessments and to review the highlights of graded assessments. Point totals and grade percentages for all assessments are provided in the table shown below.

Activity	Total Points	Final Grade Percentage
Quizzes (12 required; lowest 4 dropped)	32	16%
Assignments (5)	48	24%
Project Proposal (1)	30	15%
Project Presentation (1)	40	20%
Project Paper (1)	50	25%
Total	200	100%

Quizzes are designed to test comprehension of foundational concepts introduced in lecture videos. There are 12 **required** quizzes (the lowest four will be dropped) comprising 16% of the final grade.

Assignments are designed to keep you on track with your projects, pitches, proposals, presentations, and final papers. Students are expected to leverage assignments to explore advanced concepts and techniques, including syntactic and semantic processing using symbolic techniques, machine learning, and generative models. There are five assignments, comprising 24% of the final grade.

Both **quizzes** and **assignments** are electronically administered with a combination of automatic and hand-grading. Access to resources specified in quizzes and assignments is strongly encouraged. The details of assignment submissions will be given within each assignment specification.

NO COLLABORATION: Although you may access resources specified in quizzes and assignments, **your own work must be your own** (you may not share with others, nor may others share with you), unless explicitly indicated otherwise. **THIS COURSE HAS A ZERO TOLERANCE POLICY TO CHEATING: Cheaters will be reported to the Dean of Students Office (DSO) and will get a failing grade in the course.**

Project Pitch, Proposal, Presentation, Paper: Teams of 3-5 students will conduct a self-directed exploration of advanced concepts and will design and/or implement a proposal for further exploration, coupled with implementation and/or experimentation. The proposal (and class presentation of the proposal) is 15% of the final grade. The project (and class presentation of the project) is 20% of the final grade. Each team will produce a final paper with conclusions about the capabilities and limitations of applications and ideas for further research in the area. The final paper is worth 25% of the final grade.

Throughout the process of designing and implementing projects, students are expected to describe and provide notes on literature (e.g., from ACL, EMNLP, or AAAI conference proceedings) related to their chosen application area. These notes can be integrated into the final paper, serving as a component of the 25% final grade for the project paper.

Grading Policy

The range used to calculate your final letter grade in our course will be no harsher than the grading scale provided in the table below. Your final point total will be calculated using the percentages given in the *Evaluation of Grades* section. The percent you earn on each activity will be multiplied by the grade points associated with that activity. More information on UF grading policy may be found at: <https://catalog.ufl.edu/UGRD/academic-regulations/grades-grading-policies/>

Grade Points	Letter Grade	Highest	Lowest
4.00	A	100+	92.00
3.67	A-	91.99	89.00
3.33	B+	88.99	86.00
3.00	B	85.99	82.00
2.67	B-	81.99	79.00
2.33	C+	78.99	76.00
2.00	C	75.99	72.00
1.67	C-	71.99	69.00
1.33	D+	68.99	66.00
1.00	D	65.99	62.00
0.67	D-	61.99	59.00
0.00	E	58.99	0.00

Re-Grades

Quizzes and assignments are electronically administered, with a complete record of all attempts/saves, final answers, and the automatic grades where applicable. If you believe there is an error in a question or its possible answer or you notice a technology issue, it is important to bring this to the attention of the instructor or TA so that it can be addressed. A question can be raised in class once an assessment is completed and discussed openly in the class period. The instructor/TA will address this in a timely fashion. Re-grades typically impact the entire class. If an adjustment is made, it will be applied electronically in one fell swoop for the entire class.

More information on UF grading policy may be found at:

[UF Graduate Catalog](#)
[Grades and Grading Policies](#)

Right to change information

Although every effort has been made to be complete and accurate, unforeseen circumstances arising during the semester may require adjustments. Consequently, given due notice to students, the instructor reserves the right to change any information on this syllabus or in other course materials.

Students Requiring Accommodations

Students with disabilities who experience learning barriers and would like to request academic accommodations should connect with the disability Resource Center by visiting <https://disability.ufl.edu/students/get-started/>. It is important for students to share their accommodation letter with their instructor and discuss their access needs, as early as possible in the semester.

Course Evaluation

Students are expected to provide professional and respectful feedback on the quality of instruction in this course by completing course evaluations online via GatorEvals. Guidance on how to give feedback in a professional and respectful manner is available at <https://gatorevals.aa.ufl.edu/students/>. Students will be notified when the evaluation period opens, and can complete evaluations through the email they receive from GatorEvals, in their Canvas course menu under GatorEvals, or via <https://ufl.bluera.com/ufl/>. Summaries of course evaluation results are available to students at <https://gatorevals.aa.ufl.edu/public-results/>.

In-Class Recording

Students are allowed to record video or audio of class lectures. However, the purposes for which these recordings may be used are strictly controlled. The only allowable purposes are (1) for personal educational use, (2) in connection with a complaint to the university, or (3) as evidence in, or in preparation for, a criminal or civil proceeding. All other purposes are prohibited. Specifically, students may not publish recorded lectures without the written consent of the instructor.

A “class lecture” is an educational presentation intended to inform or teach enrolled students about a particular subject, including any instructor-led discussions that form part of the presentation, and delivered by any instructor hired or appointed by the University, or by a guest instructor, as part of a University of Florida course. A class lecture does not include lab sessions, student presentations, clinical presentations such as patient history, academic exercises involving solely student participation, assessments (quizzes, tests, exams), field trips, private conversations between students in the class or between a student and the faculty or lecturer during a class session.

Publication without permission of the instructor is prohibited. To “publish” means to share, transmit, circulate, distribute, or provide access to a recording, regardless of format or medium, to another person (or persons), including but not limited to another student within the same class section. Additionally, a recording, or transcript of a recording, is considered published if it is posted on or uploaded to, in whole or in part, any media platform, including but not limited to social media, book, magazine, newspaper, leaflet, or third party note/tutoring services. A student who publishes a recording without written consent may be subject to a civil cause of action instituted by a person injured by the publication and/or discipline under UF Regulation 4.040 Student Honor Code and Student Conduct Code.

University Honesty Policy

UF students are bound by The Honor Pledge which states, “We, the members of the University of Florida community, pledge to hold ourselves and our peers to the highest standards of honor and integrity by abiding by the Honor Code. On all work submitted for credit by students at the University of Florida, the following pledge is either required or implied: “On my honor, I have neither given nor received unauthorized aid in doing this assignment.” The Honor Code (<https://sccr.dso.ufl.edu/process/student-conduct-code/>) specifies a number of behaviors that are in violation of this code and the possible sanctions. Furthermore, you are obligated to report any condition that facilitates academic misconduct to appropriate personnel. If you have any questions or concerns, please consult with the instructor or TAs in this class.

Commitment to a Safe and Inclusive Learning Environment

The Herbert Wertheim College of Engineering values broad diversity within our community and is committed to individual and group empowerment, inclusion, and the elimination of discrimination. It is expected that every person in this class will treat one another with dignity and respect regardless of gender, sexuality, disability, age, socioeconomic status, ethnicity, race, and culture.

If you feel like your performance in class is being impacted by discrimination or harassment of any kind, please contact your instructor or any of the following:

- Your academic advisor or Graduate Program Coordinator

- Jennifer Nappo, Director of Human Resources, 352-392-0904, jpennacc@ufl.edu
- Curtis Taylor, Associate Dean of Student Affairs, 352-392-2177, taylor@eng.ufl.edu
- Toshikazu Nishida, Associate Dean of Academic Affairs, 352-392-0943, nishida@eng.ufl.edu

Software Use

All faculty, staff, and students of the University are required and expected to obey the laws and legal agreements governing software use. Failure to do so can lead to monetary damages and/or criminal penalties for the individual violator. Because such violations are also against University policies and rules, disciplinary action will be taken as appropriate. We, the members of the University of Florida community, pledge to uphold ourselves and our peers to the highest standards of honesty and integrity.

Student Privacy

There are federal laws protecting your privacy with regards to grades earned in courses and on individual assignments. For more information, please see: <https://registrar.ufl.edu/ferpa.html>

Campus Resources:

Health and Wellness

U Matter, We Care:

Your well-being is important to the University of Florida. The U Matter, We Care initiative is committed to creating a culture of care on our campus by encouraging members of our community to look out for one another and to reach out for help if a member of our community is in need. If you or a friend is in distress, please contact umatter@ufl.edu so that the U Matter, We Care Team can reach out to the student in distress. A nighttime and weekend crisis counselor is available by phone at 352-392-1575. The U Matter, We Care Team can help connect students to the many other helping resources available including, but not limited to, Victim Advocates, Housing staff, and the Counseling and Wellness Center. Please remember that asking for help is a sign of strength. In case of emergency, call 9-1-1.

Counseling and Wellness Center: <https://counseling.ufl.edu>, and 392-1575; and the University Police Department: 392-1111 or 9-1-1 for emergencies.

Sexual Discrimination, Harassment, Assault, or Violence

If you or a friend has been subjected to sexual discrimination, sexual harassment, sexual assault, or violence contact the [Office of Title IX Compliance](#), located at Yon Hall Room 427, 1908 Stadium Road, (352) 273-1094, title-ix@ufl.edu

Sexual Assault Recovery Services (SARS)

Student Health Care Center, 392-1161.

University Police Department at 392-1111 (or 9-1-1 for emergencies), or

<http://www.police.ufl.edu/>.

Academic Resources

E-learning technical support, 352-392-4357 (select option 2) or e-mail to Learning-support@ufl.edu. <https://lss.at.ufl.edu/help.shtml>.

Career Connections Center, Reitz Union, 392-1601. Career assistance and counseling; <https://career.ufl.edu>.

Library Support, <http://cms.uflib.ufl.edu/ask>. Various ways to receive assistance with respect to using the libraries or finding resources.

Teaching Center, Broward Hall, 392-2010 or 392-6420. General study skills and tutoring.
<https://teachingcenter.ufl.edu/>.

Writing Studio, 302 Tigert Hall, 846-1138. Help brainstorming, formatting, and writing papers.
<https://writing.ufl.edu/writing-studio/>.

Student Complaints Campus: <https://sccr.dso.ufl.edu/policies/student-honor-code-student-conduct-code/>; <https://care.dso.ufl.edu>.

On-Line Students Complaints: <https://distance.ufl.edu/state-authorization-status/#student-complaint>.