Natural Language Processing
CAI 6307

Section: 1CAI (26564) [F2F Only]
Class Periods: T Periods 9-11 (4:05 PM – 7:05 PM)
Location: LIT 0101
Academic Term: Fall, 2025
Instructor: Bonnie J Dorr (contact through Canvas only; include Lead TA)

Teaching Assistants: Please contact through Canvas or Slack (include all TAs)
Shlok Gilda (**Lead TA** and mentor for 6307)
Chathuri Jayaweera (Co-Lead TA for 4641; 6307 mentor)
Sangpil Youm (Co-Lead TA for 4641; 6307 mentor)

Office Hours: Classroom time is structured to provide ample opportunity for open interaction and frequent, in-person mentoring on research projects, presentations, and a final paper.

Course Description
Students will learn about Natural Language Processing research and theoretical foundations, leveraging concepts from basic bag-of-words to advanced representations, applied to tasks like translation, extraction, and dialogue. It explores both classic and cutting-edge techniques, addressing challenges and recent proposals using symbolic, machine learning, generative models, and evaluation techniques. This is a team-driven NLP project course, designed for:

1. **PhD students** pursuing AI research toward a dissertation and/or publication(s).
2. **MS** pursuing AI research toward a publication and/or thesis and have a pending application (or acceptance) to at least one PhD program.
3. **Fourth-year** and **Fifth-year** undergraduates who have introductory NLP training (e.g., CAP 4641) and/or are pursuing AI research toward an honors thesis, senior design project, or publication(s).

Course Pre-Requisites / Co-Requisites
Prerequisite: 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 **Module 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** coupled with in-class **discussion and mentoring sessions**.
- 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 NLP topics/applications, including:
    - Oral delivery of a **research project proposal** and **final research project.**
    - Oral **response to red teaming questions.**
    - 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.

Relation to Program Outcomes (ABET):

Outcome	Coverage*
1. An ability to identify, formulate, and solve complex engineering problems by applying principles of engineering, science, and mathematics	High
2. An ability to apply engineering design to produce solutions that meet specified needs with consideration of public health, safety, and welfare, as well as global, cultural, social, environmental, and economic factors	High
3. An ability to communicate effectively with a range of audiences	
4. An ability to recognize ethical and professional responsibilities in engineering situations and make informed judgments, which must consider the impact of engineering solutions in global, economic, environmental, and societal contexts	Medium
5. An ability to function effectively on a team whose members together provide leadership, create a collaborative and inclusive environment, establish goals, plan tasks, and meet objectives	
6. An ability to develop and conduct appropriate experimentation, analyze and interpret data, and use engineering judgment to draw conclusions	
7. An ability to acquire and apply new knowledge as needed, using appropriate learning strategies	High

**Coverage is given as high, medium, or low. An empty box indicates that this outcome is not covered or assessed in the 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: Jan 2025 (updated regularly at https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf)
- Textbook is referred to as "SLP" in the Course Schedule below.
- *Note: One reading from slp2 on "Chunking" will be provided in Module 6 (see Canvas)*

Supplemental/Recommended Materials

Recommended online, selected readings (freely available):

- Proceedings of ACL, EMNLP, NAACL, for example: https://aclanthology.org/events/acl-2025/
- 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/
- A primer on spaCy tool for NLP is published here: https://course.spacy.io/en
- NOTE: Optional Crossover CAP 4641 Projects have additional readings.

Required Computer

UF student computing requirement:
https://news.it.ufl.edu/education/student-computing-requirements-for-uf/

Course Schedule: Below is a general schedule for the semester. Please refer to Canvas *Announcements* and *Modules* for final details about each topic, exact dates, and this semester's schedule.

Module	Topic	Activity & Scaffolding	Assessment *Due Date*	Extra Credit Video Topics
0	Orientation	Review Syllabus, complete all items in this module. *Research Asst 1-5 out*	**Module 0 Videos, Orientation Quiz, HPG Quiz** ***Aug 25, 11:59pm***	Orientation: What is NLP? What are LLMs?
1	Research Project Planning & Team Formation	F2F 1, Tues 8/26 *All Modules Unlocked, including CAP 4641 Cross-over projects*		Language & Intelligence, NLP Pipeline, Ambiguity, Applications
2	Research Project Planning & Team Finalization	F2F 2, Tues 9/2		Text Processing in Python, Corpus Analysis
3	Research Proposal Planning	F2F 3, Tues 9/9 *1-1 meeting with team mentor*	**Asst 1** ***Sept 14, 11:59pm***	N-gram Language Models
4	Proposal Presentation Planning	F2F 4, Tues 9/16		Context Free Grammars (CFGs), Constituency Trees, CFG Parsing
5	Proposal Slide Preparation	F2F 5, Tues 9/23 *1-1 meeting with team mentor*	**Asst 2** ***Sept 25, 11:59pm*** **Proposal Slides to red-team** ***Sept 26, 11:59pm*** **Red-team Q** ***Sept 29, 11:59pm***	Grammar Conversion, CKY, Dependency Parsing
6	Proposal Presentations and Red Teaming	F2F 6, Tues 9/30	**Asst 3** ***Oct 5, 11:59pm***	Part of Speech (POS) & Named Entity (NE) Tagging, Sequence Labeling
7	Planning for Final Presentation / Paper	F2F 7, Tues 10/7		Hidden Markov Models (HMMs), Viterbi, Conditional Random Fields (CRFs)
8	Planning for Final Presentation / Paper (cont)	F2F 8, Tues 10/14 *1-1 meeting with team mentor*		Lexical Semantics, Word Meaning, Vector Semantics
9	Planning for Final Presentation / Paper (cont)	F2F 9, Tues 10/21	**Asst 4** ***Oct 26, 11:59pm***	Word Similarity, Vector Embeddings, Basic Metrics
10	Planning for Final Presentation	F2F 10, Tues 10/28		Semantic Role Labeling (SRL), Research Talk

11	Planning for Final Presentation (cont)	F2F 11, Tues 11/04 *1-1 meeting with team mentor*	**Asst 5** ***Nov 9, 11:59pm***	Contextualized Embeddings, BERT
VETERAN'S DAY: 11/11				
12	Final presentation Slides to Red Team	No F2F session due to holiday. *Work on final presentations and deliver to red team*	**Final Presentation Slides to red-team** ***Nov 14, 11:59pm***	Machine Translation, Divergences, Encoder-Decoders, Gen AI Research Talk
13	Final Presentations and Red Teaming	F2F 12, Tues 11/18	**Red-team Q** ***Nov 17, 11:59pm***	NLP Eval Methodologies, Metrics
THANKSGIVING: 11/24-11/28				
14	Final Paper	F2F 13, Tues 12/02 *Use Class time to Finalize Paper*	**Final Research Paper** ***Dec 3, 11:59pm***	*Consider input from final presentations and red teaming to finish paper.*

Format and Expectations, Attendance and Makeup Policy, How to Learn Anew or Refresh

Class time occurs in our assigned face-to-face classroom [F2F]: **Tues 4:05 PM – 7:05 PM EST (Periods 9-11)**. If you're feeling unwell, please prioritize your recovery and stay home. (More on this point under "General Attendance" below.)

Format: F2F sessions include (a) discussion of NLP problems for team projects; (b) proposal presentations/critiques; (c) presentations/critiques of hypotheses, findings, outcomes, limitations, and conclusions; and (d) one-on-one team–mentor meetings. Attend class to benefit from in-person mentoring and discussions. Outside of class, use the #general Slack channel for course-related peer discussions. If you miss a class, do not email the professor or TAs—watch the F2F recordings and/or get notes from your team (see "General Attendance" below).

For archival and review purposes, our F2F class sessions will be recorded. Notice: Any verbal participation (e.g., asking or answering questions) may be captured in the class recording. During peer presentations, make a concerted effort to demonstrate mutual respect by supporting others as you would during your own presentation. If you wish to comment or ask a question, please raise your hand to be recognized.

General Attendance: This is a face-to-face (F2F) course requiring in-person attendance. Minimally, attendance is **mandatory** during research project presentations and one-one mentoring sessions (modulo necessary absence, e.g., due to illness). If you're unable to participate in your team's presentation or mentoring session, please notify the TAs and make every effort to join using a team member's laptop or cellphone connection in a mentoring session. If electronic means are not possible, please ensure that another team member is assigned as your proxy to cover your portion of the presentation or to contribute to the mentoring session on your behalf.

Excused absences: In this course, exceptions for missed graded activities are rare and apply only to events you could not have anticipated before or during the multi-day window of the activity (projects, presentations, and assignments). The Dean of Students Office (www.dso.ufl.edu) provides rules, guidance, and approval for excuse documentation. An instructor notification letter from the DSO must directly state that an "**absence has been excused**" and the letter must specify the dates of the entire

multi-day duration of the assessment. If the letter does not meet these requirements, the student must provide the instructor with documentation that complies with UF's official excused absence policy. (See https://catalog.ufl.edu/UGRD/academic-regulations/attendance-policies/#absencestext.) Without this, the absence will not be considered excused, and accommodations cannot be provided.

Makeups or extensions: Makeups or extensions for graded class activities are provided given appropriate documentation is presented in a timely fashion (see above). However, flexibility is already built into the schedule to avoid the need for makeups or extensions. For example, required research assignments may be done at a time you select within a multi-day window. This usually avoids the need for makeups or extensions and lets you plan ahead for a timeframe that fits scheduling constraints.

Late Policy for research assignments: If you submit one day late, there will be a penalty of 50% on the graded work. For example, an assignment that is due on a Sunday at 11:59pm is **late** starting after 11:59pm Monday (even if just one second late), and will be graded 50% off up until 11:59pm on Monday. No credit is awarded starting after 11:59pm Monday.

Evaluation of Grades

Assessments are scheduled within non-overlapping multi-day time windows. This format provides ample opportunities to interactively prepare for upcoming assignments, presentations, and your final paper. Below is the breakdown of total points and their contribution to the final grade. Note: Research Assignments, Project Proposal, Project Presentation, and Project Paper are electronically administered with a combination of automatic and hand-grading.

Activity	Total Points	Final Grade Percentage
Research Assignments (5)	320	40%
Project Proposal (1)	140	17.5%
Project Presentation (1)	140	17.5%
Project Paper (1)	200	25%
Total	**800**	**100%**

Learn Anew or Refresh: If you are new to NLP or would like a review, watch our **recently updated, professionally pre-recorded lecture videos.** These Extra Credit "CAP 4641 crossover" videos (refreshed summer 2025) capture foundational NLP material and are available for viewing outside of class. Each module has 2-4 video lectures for a total of 40 videos. Begin with the Orientation (Module 0) videos before the first class. Completing Module 0 unlocks the remaining modules, which you may view at your own pace and in any order, though sequential viewing is recommended to stay on track. Video lectures include PlayPosit "thought questions" to reinforce core concepts. You may attempt these questions multiple times. Videos in Modules 1-13 may be used to earn extra credit (see later section). The textbook provides additional background for these videos, as outlined in this table.

Instructional Materials for CAP 4641 Cross-Over Videos
Module 0: SLP 1 intro, 1.1-1.4(draft)
Module 1: SLP 2 intro, 2.1-2.2; 18 intro, 18.1; 19 intro, 19.1, 19.1.1
Module 2: SLP 2.3–2.7
Module 3: SLP 3.1–3.5
Module 4: SLP 18.2, 18.5
Module 5: SLP 18.4, 18.6, 19.2
Module 6: SLP 17.1–17.3, 18.3, SLP2 13.5 (
Module 7: SLP 17.4–17.5
Module 8: SLP 6.1–6.3
Module 9: SLP 6.4–6.8, SLP 4.7
Module 10: SLP 21.1-21.6
Module 11: SLP 9.1, 10.1, 10.3, 11.1–11.3
Module 12-13: SLP 13.1–13.3, 13.6.1–13.6.2

Research Assignments: These assignments help students stay on track with projects, proposals, presentations, and final papers. Using materials from F2F sessions and Canvas modules, students will explore advanced concepts and techniques, including (but not limited to) syntactic and semantic

processing with symbolic methods, machine learning, and generative models. **Collaboration may be allowed for certain portions of the Research Assignments—but not all**. Any permitted collaboration will be explicitly stated. If not otherwise indicated, assume all work must be your own, and ask for clarification if needed. Even when collaboration is allowed, you must submit research assignments independently. Five research project assignments total **40% of your final grade**.

Project Proposal, Presentation, Paper: These collaborative assignments require teams of 4–5 students to explore advanced concepts, develop a proposal, and conduct implementation and/or experimentation. The proposal and presentation contribute 17.5% to the final grade, the project and presentation 17.5%, and the final paper 25%. The final paper will analyze the application's capabilities and limitations and suggest future research directions. Throughout the process of designing and implementing projects, students are expected to describe and provide notes on related literature (e.g., from ACL, EMNLP, or AAAI conference proceedings). These notes can be integrated into the final paper, serving as a component of the 25% final grade for the project paper.

Extra Credit: You may earn up to 2 extra credit percentage points (for a max final score of 102%) by completing the PlayPosit "thought questions" embedded in the lecture videos. The course includes 40 videos across Modules 0-13. Module 0, which is **required** and not eligible for extra credit, has two videos without thought questions. The remaining 38 videos **(Modules 1-13)** each contain a varying number of thought questions, and your performance on these determines how much of the 2 extra credit points you earn. You may attempt each video and its thought questions as many times as you like. Beyond extra credit, these videos are a key resource for your team project: earlier videos lay the groundwork for selecting topical areas around which you will develop your project, while later ones help you explore application areas and evaluation metrics that may also be relevant to your project. **So please do have a look, especially if you are new to NLP.**

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

All assessments are electronic, with a complete record of all attempts/saves and final answers. 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 professor or TA so that it can be addressed. The professor/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.

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.

The UF grading policy is at: https://catalog.ufl.edu/ugrad/current/regulations/info/grades.aspx

Visit https://go.ufl.edu/syllabuspolicies for Academic policies and resources—including student accommodations—as well as Campus Health and Wellness Resources.