Syllabus
CSCI 497.01B – Natural Language Processing
Texas A&M University Commerce
Spring 2024

Instructor: Dr. Omar El Ariss
Office Location: JOUR 238
Email: Omar.El.Ariss@tamuc.edu
Phone: 903-886-5403
Communication Response Time: 24 hours
(Please send a second email if you did not receive a response after 2 days)
Office Hours: Online through Zoom / On-campus meeting

Day	Time
TR	3:30 pm - 5:00 pm
W (Zoom)	4:30 pm - 6:00 pm

There are many ways to reach me. There is no substitute for face-to-face communication which often leads to more refined and focused questions resulting in your improved understanding. I strongly encourage you to take advantage of my office hours. Email is an easy way to ask questions outside of class but is not productive as face-to-face communication. We can also use Zoom or Skype for an online face to face communication.

Meeting Time and Place

Tuesday, Thursday: 12:20 pm to 1:35 pm, JOU 200

Recommended Textbooks

- *Speech and Language Processing*, by Daniel Jurafsky and James H. Martin, 3rd Edition (Draft), 2023.
- *Natural Language Processing with Transformers, Revised Edition*, by Lewis Tunstall, Leandro von Werra, Thomas Wolf, O'Reilly Media, 2022.

Prerequisite

- COSC 1437 (Min Grade C)

Course Objectives

Natural language processing is a crucial field of artificial intelligence, with various cutting-edge real-world applications thanks to the advances of deep learning. This course presents an overview of the theories and practices of Natural Language Processing. It covers different concepts and techniques on how to build software that works with natural language. Topics include, syntax, semantics discourse; and language models. Students will also be introduced to various applications of natural language processing such as spell checking, question answering, conversational systems, and sentiment analysis.

The main objective of this course is to learn and apply various algorithms and methods for building applications to process human language. This course will provide theoretical and practical knowledge for applications that

learn, interpret, and/or generate natural language. Upon completion of the course, students will have a good understanding of and appreciation for natural language processing and be capable of planning and applying various techniques and computational tools to build natural language processing models.

Topics:

Topics to be covered (as time permits):

- Foundations:
  - Data cleaning & preprocessing
  - Regular expressions & finite state machines
  - Lexical analysis
  - Syntax & Parsing
  - Language Models
  - Transformers
- Applications:
  - Part-of-speech tagging
  - Text Classification
  - Text summarization and generation
  - Word Sense Disambiguation
  - Question Answering

Course Outcomes

Upon completion of this course:

1) Students will gain knowledge of basic NLP tasks like tokenization, stemming, lemmatization, part of speech tagging, named entity recognition, and sentiment analysis.
2) Students will be able to implement NLP applications using Python libraries such as NLTK, spaCy, and Hugging Face.
3) Students will gain hands-on experience needed to apply NLP techniques to real-world problems.
4) Students will gain experience working with large collection of language datasets.

Assignments & Project

There will be a number of programming assignments. In addition, each student will work on a programming project on an NLP related problem, and then present it to the class. More details about the project and its requirements will be provided later.

Texas A&M University-Commerce acknowledges that there are legitimate uses of Artificial Intelligence, chatbots, or other software that has the capacity to generate code, and textual answers. Any use of such software is not allowed, and constitutes an instance of academic dishonesty (plagiarism).

Grading

- Assignments: 35% of grade
- Project: 15% of grade
- Midterm Exam: 15% of grade
- Final Exam: 35% of grade

Letter grades will be determined using a standard percentage of points scale:

Letter Grade	Cut-off Score
A	90%
B	80%
C	70%
D	60%
F	Below 60%

Doing all your assignments and project will help the borderline cases. Check your grades often. Any score may be disputed up to seven (7) days after the score is posted. After 7 days the score remains as-is.

Technology Requirements - LMS

All course sections offered by Texas A&M University-Commerce have a corresponding course shell in the myLeo Online Learning Management System (LMS). Below are technical requirements
LMS Requirements:
https://community.brightspace.com/s/article/Brightspace-Platform-Requirements

LMS Browser Support:
https://documentation.brightspace.com/EN/brightspace/requirements/all/browser_support.htm

YouSeeU Virtual Classroom Requirements:
https://support.youseeu.com/hc/en-us/articles/115007031107-Basic-System-Requirements

Access and Navigation

You will need your campus-wide ID (CWID) and password to log into the course. If you do not know your CWID or have forgotten your password, contact the Center for IT Excellence (CITE) at 903.468.6000 or helpdesk@tamuc.edu.

Note: Personal computer and internet connection problems do not excuse the requirement to complete all course work in a timely and satisfactory manner. Each student needs to have a backup method to deal with these inevitable problems. These methods might include the availability of a backup PC at home or work, the temporary use of a computer at a friend's home, the local library, office service companies, Starbucks, a TAMUC campus open computer lab, etc.

Communication and Support

If you have any questions or are having difficulties with the course material, please contact your Instructor.

Technical Support

If you are having technical difficulty with any part of Brightspace, please contact Brightspace Technical Support at 1-877-325-7778. Other support options can be found here:
https://community.brightspace.com/support/s/contactsupport

Methods of Instruction

The course will consist mainly of live lectures, discussions and student presentations. Important material from the text and outside sources will be covered during the lecture. Therefore, attending the lectures are essential for success. Students are expected to contribute each week by attending one of the office hours and asking questions.

This syllabus contains an overview of what will be covered in class; for specific information, students are referred to the class web page maintained on D2L course management system. The course web page will contain lectures, assignments, project information and supporting material. Information on D2L will be updated frequently so it is a good idea to check it regularly. Assignments are posted on D2L and should be submitted through D2L.

Late Submissions Policy

All work submitted electronically must be submitted by midnight of the due date. Late work will be deducted 10% for each day past the due date. Assignment will not be accepted after three days from the due date.

Tips for Success in the Course

- Read all assigned textbook and supplemental materials.
- Check D2L at least once a day.
- Read the textbook before and after every lecture, and use the provided lectures as guidelines.
- Practice the examples and practice exercise we go through during the lectures.
- Start your assignment & project early.
- Do your own work. Please do not copy other's work.
- Contact the instructor if you have difficulties in lecture material and/or the assignments.

Make-up Policy

No individual make-up test will be permitted except in the case of a formal institutional excuse. There will be no makeup for project deliverables.

Syllabus Change Policy

The syllabus is a guide. Circumstances and events, such as student progress, may make it necessary for the instructor to modify the syllabus during the semester. Any changes made to the syllabus will be announced in advance.

Tentative Course Outline

Week	Content
1	Introduction & Syllabus
2	Python Review, Intro to Pandas & NumPy
3	Regular Expressions & Finite State Machines
4	Lexical Analysis
5	Syntax & Parsing
6	Language Models
7	Transformers
8	Intro to NLTK
9	**Midterm Exam**, Intro to spaCy
10	**Spring Break**
11	Intro to spaCy
12	NLP Applications with Python Libraries
13	NLP Applications with Python Libraries
14	NLP Applications with Python Libraries
15	NLP Applications with Python Libraries

16	Project Presentations
17	Project Presentations
18	**Final Exam** (comprehensive)

The course outline will adapt to the actual progress of the classes and may not be accurately the same as the table above.

Student Conduct

All students enrolled at the University shall follow the tenets of common decency and acceptable behavior conducive to a positive learning environment. The Code of Student Conduct is described in detail in the Student Guidebook.
http://www.tamuc.edu/Admissions/oneStopShop/undergraduateAdmissions/studentGuidebook.aspx

Students should also consult the Rules of Netiquette for more information regarding how to interact with students in an online forum: https://www.britannica.com/topic/netiquette

Academic Honesty

"All students enrolled at the University shall follow the tenets of common decency and acceptable behavior conducive to a positive learning environment." (See Student's Guide Handbook, Policies and Procedures, Conduct). It is the policy of the University, that no form of plagiarism or cheating will be tolerated. Plagiarism is defined as the deliberate use of another's work and claiming it as one's own. This means ideas as well as text or code, whether paraphrased or presented verbatim (word-for-word). Cheating is defined as obtaining unauthorized assistance on any assignment. Proper citation of sources must always be utilized thoroughly and accurately. If you are caught sharing or using other people's work in this class, you will receive a 0 grade and a warning on the first instance. A subsequent instance will result in receiving an F grade for the course, and possible disciplinary proceedings. If you are unclear about what constitutes academic dishonesty, ask.
For more details and the definition of academic dishonesty see the following procedures:

Graduate Academic Dishonesty 13.99.99.R0.10

Special Needs

The Americans with Disabilities Act (ADA) is a federal anti-discrimination statute that provides comprehensive civil rights protection for persons with disabilities. Among other things, this legislation requires that all students with disabilities be guaranteed a learning environment that provides for reasonable accommodation of their disabilities. If you have a disability requiring an accommodation, please contact:

Office of Student Disability Resources and Services
Texas A&M University-Commerce
Gee Library- Room 132
Phone (903) 886-5150 or (903) 886-5835
Fax (903) 468-8148
Email: studentdisabilityservices@tamuc.edu
Website: Office of Student Disability Resources and Services
http://www.tamuc.edu/campusLife/campusServices/studentDisabilityResourcesAndServices/

A&M-Commerce Supports Students' Mental Health

The Counseling Center at A&M-Commerce, located in the Halladay Building, Room 203, offers counseling services, educational programming, and connection to community resources for students. Students have 24/7

access to the Counseling Center's crisis assessment services by calling 903-886-5145. For more information regarding Counseling Center events and confidential services, please visit www.tamuc.edu/counsel

Nondiscrimination Notice

Texas A&M University-Commerce will comply in the classroom, and in online courses, with all federal and state laws prohibiting discrimination and related retaliation on the basis of race, color, religion, sex, national origin, disability, age, genetic information or veteran status. Further, an environment free from discrimination on the basis of sexual orientation, gender identity, or gender expression will be maintained.

Campus Concealed Carry Statement

Texas Senate Bill - 11 (Government Code 411.2031, et al.) authorizes the carrying of a concealed handgun in Texas A&M University-Commerce buildings only by persons who have been issued and are in possession of a Texas License to Carry a Handgun. Qualified law enforcement officers or those who are otherwise authorized to carry a concealed handgun in the State of Texas are also permitted to do so. Pursuant to Penal Code (PC) 46.035 and A&M-Commerce Rule 34.06.02.R1, license holders may not carry a concealed handgun in restricted locations.

For a list of locations, please refer to the Carrying Concealed Handguns On Campus document and/or consult your event organizer.

Web url:
http://www.tamuc.edu/aboutUs/policiesProceduresStandardsStatements/rulesProcedures/34SafetyOfEmployees AndStudents/34.06.02.R1.pdf

Pursuant to PC 46.035, the open carrying of handguns is prohibited on all A&M-Commerce campuses. Report violations to the University Police Department at 903-886-5868 or 9-1-1.