

Learning Outcomes:

By the end of this course all students will be able to:

- Acquire and pre-process textual data.
- Apply Machine Learning algorithms to different NLP tasks.
- Evaluate and analyze the outputs of NLP models.

By the end of this course graduate students will also be able to:

- Implement end-to-end NLP solutions for real world problems.

ISTA Course Competencies:

F1 (Introduced & Practiced)

Students will be able to perform the 4 fundamental tasks of research (design, data collection, analysis, and presentation of results).

IS1 (Introduced, Practiced & Assessed)

Students will be able to solve information and data challenges using key techniques of computational thinking (decomposition, pattern recognition, abstraction, and algorithms).

IS2 (Introduced, Practiced & Assessed)

Students will be able to collect, manipulate and correctly analyze varied data at different scales.

INFO Course Competencies:

MS1 (Introduced, Practiced & Assessed)

Students will establish the ability to exercise the four key techniques of computational thinking: decomposition, pattern recognition, abstraction, and algorithms.

MS2 (Introduced, Practiced & Assessed)

Students will obtain the skills of collecting, manipulating, and analyzing different types of data at different scales, and interpreting the results properly.

Course Description

Most of the data available on the web or managed by institutions and businesses consists of unstructured text. Natural language processing tools help to organize such texts, extract relevant information from them, and automatize time-consuming processes. This course will teach the fundamental knowledge necessary to design and develop end-to-end natural language understanding applications, drawn from examples such as question answering, sentiment analysis, information extraction, automated inference, machine translation, chatbots, etc. We will use several natural language processing toolkits and libraries.

Online Contact Information

All questions to the instructor should be asked via Slack at <https://ua-info555-sp25.slack.com>. You

can add yourself the workspace here: https://join.slack.com/t/ua-info555-sp25/shared_invite/zt-30wom9yte-7gbXjc56BOQeCqt8N_eFaQ. The Slack workspace contains a channel for each module of the course and a channel for programming questions related to the assignments. For general questions about the course, students can use the *general* channel.

Response Time

The instructor will typically get back to you within 48 hours. Do not expect any reply on weekends.

Note!

You will be notified, in advance, of any scheduling issues that may impact our response times.

Instructor Information

Professor

[Dr. Egoitz Laparra](#)

Office Location

[Harvill Building, Rm 435D](#), but office hours will be conducted via Zoom

Office Hours

By appointment [via Calendly](#).

Other Contact Information

Graduate Teaching Assistant

Aslam Sheik Dawood

Office Location

Office hours will be conducted via [Zoom](#)

Office Hours

Thursday 11:00am - 1:00am Arizona time

Meeting Information

This class is scheduled to be taught in ONLINE modality.

- **Academic advising:** If you have questions about your academic progress this semester, please reach out to your academic advisor (<https://ua-trellis.my.site.com/Public/s/advisors-by-major>). Contact the Advising Resource Center (<https://advising.arizona.edu/>) for all general advising questions and referral assistance. Call 520-626-8667 or email to advising@arizona.edu

- **Life challenges:** If you are experiencing unexpected barriers to your success in your courses, please note the Dean of Students Office is a central support resource for all students and may be helpful. The [Dean of Students Office](#) can be reached at (520) 621-2057 or DOS-deanofstudents@email.arizona.edu.
- **Physical and mental-health challenges:** If you are facing physical or mental health challenges this semester, please note that Campus Health provides quality medical and mental health care. For medical appointments, call (520) 621-9202. For After Hours care, call (520) 570-7898. For the Counseling & Psych Services (CAPS) 24/7 hotline, call (520) 621-3334.
- **Class Content:** Students may not modify content or re-use content for any purpose other than personal educational reasons. All recordings are subject to government and university regulations. Therefore, students accessing unauthorized recordings or using them in a manner inconsistent with [UArizona values](#) and educational policies ([Code of Academic Integrity](#) and the [Student Code of Conduct](#)) are also subject to civil action.

Activities and Assignments

This is a *collaborative learning* class. Most units will include one or more quizzes that you will complete on your own, and then discuss in a group. Modules also contain group activities, where you will debate with your group issues related to the development of NLP projects. As a minimum requirement, you are expected to at least justify your answers to the quizzes and give your opinion on the question proposed for debate. However, you are encouraged to discuss with your group mates the different answers and points of view.

Graded Activities

Quizzes

18% (ISTA 455) or 15% (INFO 555) of the grade.

There will be 15 single-question, multiple-choice quizzes distributed throughout the course on a topic from the corresponding assigned reading. You will receive a score of 100% if you answer it correctly, 80% if you answer it incorrectly, and 0% if you do not answer.

Programming Assignments

70% (ISTA 455) or 75% (INFO 555) of the grade

There will be 4 programming assignments in the form of Jupyter Notebooks, which must be written in Python and submitted via GitHub Classroom. The objective of each assignment is to solve an NLP task by completing small programming problems. Each notebook will provide some guidance and pointers to documentation that will help you solve the coding problems. The assignments will be graded on their ability to pass the tests that are provided to you. You will receive 80% of the possible points by solving the basic problems in the assignment. The remaining 20% of the points are for solving more advanced and challenging problems. For graduate students, there will be an additional task on the last assignment. For undergraduate students each assignment will count 15%

of the grade. For graduate students, the first 3 assignments will count 15% and the final one 30% of the grade.

Final Exam

12% (ISTA 455) or 10% (INFO 555) of the grade.

There will be a final exam that will cover the content of the lectures and the reading assignments given during the course. Students may use the material provided, their own notes and other resources online. The exam will be posted to D2L and will be available for 3 days during Finals Week. Students may respond to the exam at any time while it is available. Check the course calendar for exact dates.

[Final Exam Schedule](#)

Other Activities

Readings

All readings will be provided within the course. Students will be expected to read both formal and informal articles, books, and documents.

Quiz Discussions

After every Quiz, students will be required to participate in a group discussion by justifying their answers. This justification does not have to be very long, one or two sentences are sufficient in most cases, nor does it have to be correct. An inaccurate justification will not affect the quiz grade. The correct answer for the Quiz will be unlocked after the student posts their justification.

Group Activities

Although they will not be evaluated for the final grade, students are expected to engage in group activities in the form of online discussions. Throughout the course, different questions open to debate will be proposed. The objective of the student is not to give a correct answer, but rather to give their own opinion and discuss with their classmates the different points of view. The instructor will post their own answer after all group members have participated in the discussion.

Important

Although Quiz Discussions and Group Activities are not graded, students are expected to participate fairly and honestly, avoiding cheating. Each failure in this regard may result in a penalty of minus 10 points on the final grade.

Table of Contents

The course content will be unlocked as the student completes the lectures and activities. Students

are expected to progress at their own pace, although it is recommended that they have covered the associated content by the time the assignments are published.

Module 0: Introduction to NLP projects

- Overview of ML for NLP
- Overview of the NLP workflow

Module 1: Textual Data

- Data Acquisition
- Manual Annotation
- Pre-Processing

Module 2: Text Representation

- Discrete Text Representations
- Word Embeddings
- Semantic Text Similarity

Module 3: Text Classification

- Text Classification with Linear Regression
- The FeedForward Neural Network for NLP
- Evaluation Metrics

Module 4: Sequence Processing

- Sequence Labeling
- Reading Comprehension
- Sequence to Sequence

Module 5: Pre-trained Language Models

- Language Modeling
- Pre-trained Language Models
- PLM Variants

Module 6: A Paradigm Shift

- Task Unification
- Instructed PLMs

Important

Information contained in the course syllabus, other than the grade and absence policy, may be subject to change with advance notice, as deemed appropriate by the instructor.

Texts, Articles, and Other Course Materials

This course does not follow any particular book, the class will be assigned readings from recent different sources. However, it is recommended that students read at least one book on NLP that covers ML and DL algorithms, for example:

- Daniel Jurafsky and James H. Martin. 2021. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Third Edition. <https://web.stanford.edu/~jurafsky/slp3/>
- Jacob Eisenstein. 2019. Introduction to Natural Language Processing. The MIT Pres. <https://mitpress.mit.edu/9780262042840/introduction-to-natural-language-processing/>

Grading Policies

This course follows grading policies established by the University of Arizona.

Note!

Click here: [for descriptions of grades given at the U of A and to review the official university grading system.](#)

Grading Scale

A = 90% and above

B = 80 - 89.9%

C = 70 - 79.9%

D = 60 - 69.9%

E = 59.9% and below

About Policies

Policies are a set of guiding principles for how you (the student), we (the instructors), and the university should act in a given situation. Read these policies carefully so you know what is expected of you as well as what you can expect from the course and the UA.

Course Policies

Communication Policy

All online communication for this course should occur through Slack.

Classroom Behavior Policy

To foster a positive learning environment, students and instructor 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.

Elective Name and Pronoun Usage

It is already UA policy that class rosters are provided to instructors with a student's preferred name. Students may share their preferred name and pronoun with members of the teaching staff and fellow students, as desired, and these gender identities and gender expressions will be honored in this course. As the course includes group work discussion, it is critical to create an educational environment of inclusion and mutual respect. In this class, to be inclusive of all gender identities and expressions, students will be referred to by their first or last names, the pronoun of their choice, or by default, the pronoun "they".

Late Work - Makeup Policy

Programming assignments may be turned in late, but late assignments will receive at most 70% of the maximum grade and only limited code feedback.

Academic Integrity in this course

This policy agreed upon by faculty in the UArizona iSchool applies in addition to the Dean of Students' [Code of Academic Integrity](#).

Students in courses at the UArizona iSchool are expected to maintain rigor in their academic performance with intent to learn, practice, and overcome challenges toward personal growth and enrichment. As future professionals in digital environments, iSchool students are also expected to exercise transparency and integrity in collaborations and in the use of tools and resources that may aid completion in assignments for our courses.

Consider the following PROHIBITED practices in this course, unless the instructor has *specifically written instructions or permission to do otherwise*:

- Posting a question on an online site such as Chegg.com, and copying and pasting some or all of the response into an assessment
- Posting an assessment from the course on online sharing sites such as Course Hero. Aiding other students in violation of academic integrity is also a violation, and is potential copyright infringement.
- Generating and submitting, in whole or in part, text or code through Artificial Intelligence such as ChatGPT, QuillBot, and text summarizers
- Using, in whole or in part, computer code not written by the student (for example, from another student, a book, or the internet) in an assignment or project. This includes using such code in modified or unmodified form.

- Searching for solutions to projects or assignments on the internet or through other tools, when your instructor intended for you to learn the solution through exercises (e.g. Googling for the solution to a question on an assignment).
- Simultaneously submitting the same assignment as another student enrolled into the course without prior permission from the instructor

Exceptions

In any cases in which this course requires or permits students to use practices in the list above, clear written instructions will specify the tools allowed or required, so students can be certain they are working as instructed.

Plagiarism

Students may discuss their approaches to programming assignments with each other, but should never be looking directly at each other's code. If a student uses code from an online source (e.g., Stack Overflow) for part of an assignment, the student should include a link to the source used in the comments of their code. Improper sharing of code will be reported as a violation of [the university's code of academic integrity](#).

The University Libraries also have some [excellent tips for avoiding plagiarism](#).

Selling Course Materials

Selling class notes and/or other course materials to other students or to a third party for resale is not permitted without the instructor's express written consent. Violations to this and other course rules are subject to the Code of Academic Integrity and may result in course sanctions. Additionally, students who use D2L or UA email to sell or buy these copyrighted materials are subject to Code of Conduct Violations for misuse of student email addresses. This conduct may also constitute copyright infringement.

University Policies

<https://academicaffairs.arizona.edu/syllabus-policies>