SP26-DATA-0297-1 Course Spring 2026

Deep Learning for Multimodal AI

Instructor

Shuo Zhang, Ph.D., Professor of the Practice, Data Analytics, Tufts University
Office Hours: Monday 4-5 pm, JCC 684
Email: Shuo.Zhang@tufts.edu

Support staff

Industry Advisor: Jeanie Cherng, Ph.D
TA/RA: Ishmam Khan (Ishmam.Khan@tufts.edu)

Class Time

Mondays 6-9pm Anderson Wing TTC, Room 208, Medford/Somerville campus

Course GitHub Repo

github link.

1 Syllabus

1.1 Pre-requisite

Proficient in python coding; familiarity with probability, linear algebra, and calculus, and machine learning (intro level)

1.2 Course Description

Artificial Intelligence (AI) employs a variety of modalities, including image, audio, and text to interact with the world around us. This course is designed to introduce students to the machine learning and deep learning techniques applied to data in multiple modalities. Students will gain a broad understanding of how these techniques are being applied to domain-specific problems ranging from computer vision to natural language processing to audio and music understanding. Students will gain hands-on experience implementing AI algorithms and building AI systems using popular ML tools and frameworks such as PyTorch and Tensorflow/Keras using Google Colab or the GPUs from Tufts HPC. The skills learned in this course will be valuable for career paths in various industries with a diverse product portfolio.

1.3 Textbooks

We will use a combination of online reading materials and Jupyter Notebook tutorials, as well as online versions of the updated textbooks in deep learning and NLP.

- *Dive into Deep Learning*, by Aston Zhang, Zachary C. Lipton, Mu Li, Alexander J. Smola.
- *Speech and Language Processing*, 3rd edition, Jurafsky and Martin.

1.4 Grades

Course grades will be a weighted sum of two factors: homework and coding assignments (60%), and a final project (40%).

1.5 Coding Assignments

Coding assignments will be using the tensorflow keras API. To complete a coding assignment, first code out the assignment yourself in a Jupyter notebook. Commit and push your code into your branch in the class Git repo. Then meet with your teammates to for a discussion before pushing your final code.

1.6 Final Project

Students will be divided into 1-3 person teams to propose a final project to apply machine learning and deep learning to a problem in any of the modalities covered in this class: image, audio, or text. Students will use the final four weeks of the course to conduct independent research to formulate the project and give a final presentation.

1.7 Academic Policies & Student Support

Academic Integrity Policy: Tufts holds its students strictly accountable for adherence to academic integrity. Student Resources: Religious Accommodations: Tufts University faculty, staff, and administration highly value and acknowledge the religious diversity of its student body. Accommodations for Students with Disabilities: Tufts University values the diversity of our students, staff, and faculty and recognizes the important contribution each student makes to our unique community.

2 Specifics

2.1 Homework & code reviews

When a homework is assigned on Monday, team members should take until Thursday to do the homework independently. On Friday, the team lead will call a team meeting to discuss the homework. The team lead will also collect questions from the team members. Everyone should ask at least one question.

One team will be responsible for delivering a code review for a homework on the due date (Monday in class). I will post the schedule of which team to present which homework. The code review could feature 1 or more presenters.

2.2 Reading discussions

For each module (out of the 3 modules - computer vision, NLP, and audio), there is one reading discussion in class. Each team will be responsible for discussing and presenting one of the readings in class. All team members should get together to discuss the reading in your team meeting prior to the class, and then one or more team members will present the reading in class and lead a discussion.

2.3 Final Project rules

2.3.1 General

Find a partner to form a group and define a final project and carry out a month long research and implementation. Group size can vary from 2-3 (if you feel strongly that you want to be a group of one, you need to put in a special request and get the professor's approval). The final project will be delivered in a presentation and a paper.

There are several types of final project. Such as solve a well defined CV, NLP or audio ML task using a dataset and compare several different methods; build a AI product feature using one or more modalities; exploratory analysis of text/audio/image data (see NLP4DH below); replicate a research paper; propose a new method on a task; etc.

No matter the type of project, in each project:

- You must explore the use of a novel method that was not covered in this class, and this is something that you should do research on and learn about yourself. This could also be exploring a new task that was not covered in detail in class and explore methods to be applied to that task.

- The co-authors must each contribute at least one novel method to be used in this project and final paper
- The paper should compare several methods/approaches to a simple baseline and report back the results.
- You must have a limited scope to be able to finish within about 1 month.
- Time estimation. For instance, how long do you plan to do literature review, data collection and curation, writing code, etc. This is just to make sure you have a realistic chance of finishing a project in a month. For instance, if the problem you want to solve involves a very long process of data collection, you might need to think about whether you have time to finish it.
- Think about how you will evaluate your methods. If you need annotated ground truth data, think about how you will obtain it.

2.3.2 Project proposal

Components of the project (proposal) :

- Dataset
- Data preprocessing/cleaning
- Research question
- Problem description
- Scope
- Literature review (1–2 papers per person)
- Methodology (compare several novel methods not covered in class to a simple baseline)
- Evaluation
- Error analysis
- Timeline estimation

2.3.3 Deliverables

The final project will be delivered in the form of a presentation and a paper, not code.

2.3.4 Proposal Requirements

In your proposal, you can cover some of the above aspects, but it is not necessary to include all of them at the proposal stage. The proposal should primarily focus on:

- Research question
- Problem description and motivation
- Dataset / data acquisition plan
- Scope and planned literature review

2.3.5 Grading (Total: 40 Points)

- **Paper formatting and style** (5 points)
  The main paper content (excluding references) should be a minimum of 4 pages and a maximum of 6 pages, in a two-column format. The paper should be formatted using the LaTeX or Word template available here:
  `https://2023.aclweb.org/calls/style_and_formatting/`
- **Result & discussion** (5 points)
  Is there a thorough discussion on results and observations, error analysis, etc.?
- **In-class presentation** (5 points)
  Does the in-class presentation clearly discuss the motivation, problem formulation, dataset, and proposed methods?
- **Literature review** (5 points)
  Thoroughly review and explain relevant existing work in this domain.
- **Novelty and Difficulty** (5 points)
  How novel is this idea? How challenging is this project to execute? This is correlated with the amount of work invested. For easier problems, more effort is expected (e.g., experimenting with more advanced methods). For harder problems, lower performance is acceptable if justified.
- **Experimental Validation** (5 points)
  Is there sufficient experimental validation to support the main claim or conclusion?
- **Clarity of Writing** (5 points)
  Is the paper clearly written? Can a reader easily understand what was done? Be specific and transparent about your work.
- **Technical Correctness** (5 points)
  Is the technical solution correct?

3 Course Schedule

Please see the updated course schedule in the Course Website.