![USC Viterbi School of Engineering]

ITP-459 Applied Machine Learning for Natural Language Processing and Generative AI
Units: 4
Term: Fall 2024
Prerequisite(s): ITP 359
Time: Mondays 5-8:20 PM
Location: SOS B4

Instructor: Allen Bolourchi
Office: Zoom
Office Hours: TBD 6:30 PM-7:30 PM by appointment only. Select your desired time 12 hours in advance here: https://calendly.com/allenusc
Contact Info: bolourch@usc.edu

Course Assistants:

Course Assistant	Email	Mode	Location	Day	Time Slot	Meeting Link
Rajeev Singh	rajeevdh@usc.edu	Online	TBD	TBD	-	
Rajeev Singh	rajeevdh@usc.edu	In Person	TBD	TBD	-	

Course Description

Learn the state-of-the-art technology in Artificial Intelligence including the latest AI tools and algorithms in Natural Language Processing, Generative AI, GPT/BERT models and Hugging Face. You will learn the fundamentals of how computers understand human languages and what technologies and products have been built based on NLP and Generative AI for text.

Co-Requisite(s): None
Concurrent Enrollment: None
Recommended Preparation: Google Colab

Course Notes

Grading Type: Letter grade
Lecture slides, assignments, readings, announcements, and other class information will be posted on Brightspace.

Technological Proficiency

Familiarity with Google Colab and Python is necessary. If you haven't used them, familiarize yourself with Google Colab and set it up. We will teach you the rest of the python packages in the classroom.
- Google Colab - https://colab.research.google.com/
- Python packages – Keras, PyTorch, TensorFlow, scikit-learn - https://www.tensorflow.org/

Description and Assessment of Assignments and Grading Breakdown

See the Course Schedule for more details and due dates.

Assignment	% of grade
Homework – Top 12 Scores	24
Gen AI Case Presentation	8
Final Project	25 + 4 Extra Credit
Midterm Exam	15
Final Exam	25
Class Surveys	1
Class Participation (3Qs)	3
TOTAL	**101-105**

Assignment Submission Policy

Students are responsible for completing individual assignments by the stated deadlines. Assignments must be turned into Brightspace. Assignments turned in late will have 25% of the total points deducted from the graded score for each 3-day late day. Each student gets 4 x 3-day no penalty days throughout the semester except during finals week due to grade submission deadlines. Please note that 3 days will be deducted from your allowance even if you are requesting only one day of extension. The free no penalty days are to cover all emergencies and family situations or sick days. No other allowance and extension are available outside this offering. The eligibility is conditioned on an email before the assignment deadline to the Course Assistant and the professor.

Gen AI Case Presentation

- The presentation can include one or two generative AI examples, with only 5-10 slides and a 1-page summary report.
- We only have 15 minutes: 10 minutes for the presentation and 5 minutes for Q&A.
- The generative AI examples in the presentation should be based on applications, corporate use cases, or new apps.
- The examples should be backed by well-known combinators such as YCombinator or well-known VCs in Pre-Seed, Seed, or Series A stages.
- Please exclude Photoshop-like, deepfake, and NLP applications that most people are familiar with or we covered in the class.
- The one-page report and the PDF of the presentation are due before class.
- The presentation should be in-person, with all team members participating.

Final Project

Final Project Submission Details

● **Selection Deadline:** Choose your final project topics by the announced dat at midnight from the provided list to receive 5 points towards your final project. Late submissions will be treated like late assignments. Please note the limit on the number of teams allowed to use a certain project.
● **Presentation Date:** Prepare a PowerPoint presentation and highlight how your code executes.
● **Video Requirement:** Produce a video (maximum of 10 minutes) demonstrating your code functionality and user interactions.
● **Coding Tools:** Use Anaconda and VS Code (recommended). You can use Colab and other tools you are comfortable with too and adjust the Streamlit code accordingly. The code guide is best utilized via VS Code.
● **Submission Deadline:** All project materials are due by the announced day at 6 PM, (code, presentation, and video)
● **Individual Submission Requirement:** Each group member must submit their team's work on Brighrspace under their own name to receive credit.

Final Presentation Guide
1.   Start by introducing your team. You can have one slide on your team.
2.   (1-2 slides) An overview of the industry and how the selected product is positioned. You can include a brief market analysis here (competitors and customers).
3.   (1-2 slides) Motivation - what is the motivation, the problem we are trying to solve and needs of a potential user/customer it will address.
4.   (2-4 slides) Product intro - what is your product, what are the features and how the user should use it.
5.   (3-4 slides) Coding and design - highlight the steps you followed to design the product. Show code snippets and the output. Add the architecture of your code (optional).
6.   (1 slides) Monetization - How this product can be monetized. Use your best intelligence and common sense to come up with a multiple tier pricing model.
7.   (4-12 slides) Success Metrics - if we launch this app/product, how would you measure the performance and the adoption rate? Number of app times, user has engaged with it. Number of premium users hit. Error you faced. R; Various metrics to v. User switch back to status quo
8.   (2-8 slides) Potential Risks and Guardrails - What is needed here to protect your product against misusing it.
9.   (1 slides) Features Improvement - How to enhance the product success.
10.   (1 slides) How would you enhance this product be ultimately successful with Venture Capital funds and many ML engineers and designers, what could be its ultimate expansion stage and the shape of the product? Can it be a billion-dollar unicorn sold to a big tech company such as Amazon?

Final Project Grading Policy		Points
Presentation		
1	Following the presentation guideline	2
2	Content of the presentation	2
3	Business Acumen	2
4	Slides organization	2
5	Delivery	2
Code		
6	Difficulty	4
7	Organization	2
8	Effort	7
Video		
9	Explanations	2
Extra Credit		
10	Extra credit	4
	Total	25-29

Exams

The ITP 459 midterm and final exam evaluates students' understanding and application of key concepts in text processing, model comparison, and coding. The exam is divided into two main parts. Sample questions will be provided prior to the exams.

Part 1 (closed laptop/notes) includes multiple-choice questions, concept explanation questions, and code explanations and corrections.
Part 2 (open laptop and class materials) involves code completion questions and coding questions.

Additional Policies

No make-up exams (except for documented medical or family emergencies, or religious observance) will be offered. If students will not be able to attend an exam due to an athletic game or other valid reason, then they must coordinate with the instructor before the exam is given. They may arrange to take the exam before they leave, with approved university personnel during the time they are gone, or within the week the exam is given. If students do not take an exam, they will receive a 0 for that exam.

If students need accommodations authorized by OSAS (*Office of Student Accessibility and Services*), they should notify the instructor at least two weeks before the exam. This will allow time for arrangements to be made.

Sharing of course materials outside of the learning environment

SCampus Section 11.12(B)
Distribution or use of notes or recordings based on university classes or lectures without the express permission of the instructor for purposes other than individual or group study is a violation of the USC Student Conduct Code. This includes, but is not limited to, providing materials for distribution by services publishing class notes. This restriction on unauthorized use also applies to all information, which had been distributed to students or in any way had been displayed for use in relationship to the class, whether obtained in class, via email, on the Internet or via any other media. (See Section C.1 Class Notes Policy).

Intellectual Property Policy: Codes and Ideas

All ideas discussed in the classroom or outside that are pertinent to the course, presentations made by students and the codes provided for assignments and the final project can be made available to all students for learning purposes anonymously and/or may be used or referred to outside the classroom. Students should consider the classroom a public domain and refrain from discussing confidential information from their work environment or their own startups.

If you have concerns about the IP of your idea or code, please consult with USC staff and IP attorneys.

IT Help: Viterbi IT
Hours of Service: Monday – Friday, 8:30 AM – 5:00 PM
Contact Info: DRB 205
(213) 740-0517 engrhelp@usc.edu

Course Schedule: A Weekly Breakdown

Week	Date	Topic
1	8/26	**Week 1: Course Overview, Introduction to NLP, AI, Generative AI, Concepts, Applications, Text Processing, Challenges** • Historical evolution, key concepts, applications, Text processing, challenges, limitations
	9/2	**Labor Day Holiday**
2	Rec.	**Week 2: Text Cleaning and Preprocessing, Normalization, Tokenization, Feature Extraction, Python Libraries** • Noise reduction, tokenization, embeddings • Feature extraction, Python libraries overview (NLTK, Spacy, regex)
3	9/9	**Week 3: Fundamentals of Machine Learning for NLP, Supervised Learning, Metrics, Cross-Validation** • Key algorithms, evaluation metrics, supervised and unsupervised learning • Training, validation, testing sets, overfitting, cross-validation
4	9/16	**Week 4: Neural Networks, Deep Learning in NLP, Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long short-term memory (LSTM), Gated recurrent units (GRUs)** • Activation functions, backpropagation, CNNs, deep learning • LSTM, GRU, sequence modeling, challenges
5	9/23	**Week 5: Syntax, Parsing, Word Embeddings, Part of Speech (POS) Tagging, Vector Database for Generative AI** • Syntax, dependency, constituency parsing • Word embeddings, POS tagging, vector database for generative AI
6	9/30	**Week 6: Generative AI and Transformer Models, Self-Attention, Bidirectional Encoder Representations from Transformers (BERT), Generative Pre-trained Transformer (GPT), Training, Challenges** • Transformer architecture, BERT, GPT, self-attention • Training models, challenges, solutions
7	10/7	**Week 7: Semantic Analysis, Language Models, Question Answering (QA), Knowledge Graphs** • Semantic role labeling, contextual embeddings • QA systems, generative models, knowledge graphs
8	10/14	**Week 8: Text Classification, Machine Translation, NMT, Midterm Exam** • Classification techniques, Neural Machine Translation model • Generative models for machine translation, evaluation metric
9	10/21	**Week 9: Advanced Topics in Machine Learning, Transfer Learning, Interpretability** • Deep transfer learning, data imbalance • Model interpretability, case studies
10	10/28	**Week 10: Named Entity Recognition (NER), Information Retrieval, TF-IDF, Search Engines** • Advanced NER techniques, contextual NER • Information retrieval, TF-IDF, search engines
11	11/4	**Week 11: Advanced Machine Translation (MT), Summarization (Extractive, Abstractive) and Challenges** • Advanced MT techniques, low-resource languages • Summarization methods with & without generative models, challenges, evaluation
	11/11	**Veterans Day Holiday**
12	11/18	**Week 12: Generative AI and Large Language Models (LLMs), Prompt Engineering, APIs, Ethical Considerations** • Generative models, prompt engineering, AI ethics, APIs, business use cases • LLM safety: prompt injection, leaking, jailbreaking
13	11/25	**Week 13: Training and Fine-Tuning LLMs, Hugging Face, LangChain, Retrieval Augmented Generation (RAG)** • LLM training, reinforcement learning, Human In The Loop (HITL) • ChatGPT training and LLM evaluation benchmarks, RAG, Hugging Face, LangChain
14	12/2	**Week 14: NLP and Generative AI Course Review, Future Trends, Final Exam** • Recap key concepts, future trends, course review, Final exam
	12/9	Study Day
15	12/16	**Week 15: Generative AI Capstone Project Presentation** • Student presentations, feedback and review

Statement on Academic Conduct and Support Systems

Academic Conduct:

Plagiarism – presenting someone else's ideas as your own, either verbatim or recast in your own words – is a serious academic offense with serious consequences. Please familiarize yourself with the discussion of plagiarism in SCampus in Part B, Section 11, "Behavior Violating University Standards" policy.usc.edu/scampus-part-b. Other forms of academic dishonesty are equally unacceptable. See additional information in SCampus and university policies on scientific misconduct, policy.usc.edu/scientific-misconduct.

Support Systems:

Counseling and Mental Health - (213) 740-9355 – 24/7 on call
studenthealth.usc.edu/counseling
Free and confidential mental health treatment for students, including short-term psychotherapy, group counseling, stress fitness workshops, and crisis intervention.

National Suicide Prevention Lifeline - 1 (800) 273-8255 – 24/7 on call
suicidepreventionlifeline.org
Free and confidential emotional support to people in suicidal crisis or emotional distress 24 hours a day, 7 days a week.

Relationship and Sexual Violence Prevention Services (RSVP) - (213) 740-9355(WELL), press "0" after hours – 24/7 on call
studenthealth.usc.edu/sexual-assault
Free and confidential therapy services, workshops, and training for situations related to gender-based harm.

Office of Equity and Diversity (OED) - (213) 740-5086 | Title IX – (213) 821-8298
equity.usc.edu, titleix.usc.edu
Information about how to get help or help someone affected by harassment or discrimination, rights of protected classes, reporting options, and additional resources for students, faculty, staff, visitors, and applicants.

Reporting Incidents of Bias or Harassment - (213) 740-5086 or (213) 821-8298
usc-advocate.symplicity.com/care_report
Avenue to report incidents of bias, hate crimes, and microaggressions to the Office of Equity and Diversity |Title IX for appropriate investigation, supportive measures, and response.

The Office of Student Accessibility and Services - (213) 740-0776
osas.usc.edu
Support and accommodations for students with disabilities. Services include assistance in providing readers/notetakers/interpreters, special accommodations for test taking needs, assistance with architectural barriers, assistive technology, and support for individual needs.

USC Campus Support and Intervention - (213) 821-4710
campussupport.usc.edu
Assists students and families in resolving complex personal, financial, and academic issues adversely affecting their success as a student.

Diversity at USC - (213) 740-2101
diversity.usc.edu
Information on events, programs and training, the Provost's Diversity and Inclusion Council, Diversity Liaisons for each academic school, chronology, participation, and various resources for students.

USC Emergency - UPC: (213) 740-4321, HSC: (323) 442-1000 – 24/7 on call

dps.usc.edu, emergency.usc.edu

Emergency assistance and avenue to report a crime. Latest updates regarding safety, including ways in which instruction will be continued if an officially declared emergency makes travel to campus infeasible.

USC Department of Public Safety - UPC: (213) 740-6000, HSC: (323) 442-120 – 24/7 on call

dps.usc.edu

Non-emergency assistance or information.

Office of the Ombuds - (213) 821-9556 (UPC) / (323-442-0382 (HSC)

ombuds.usc.edu

A safe and confidential place to share your USC-related issues with a University Ombuds who will work with you to explore options or paths to manage your concern.