UCLA
Computer Science Department

CS 97: Introduction to Generative AI   Summer 2024

Course Description: This course is an adapted from CS 162: Natural Language Processing. Natural Language Processing (NLP) is a rapidly developing field, with recent advances of deep neural networks that revolutionize many NLP applications. This course is intended as an introduction to a wide range of NLP tasks, algorithms for effectively solving these problems (including the most recent advances of deep learning models), and methods for evaluating their performance. There will be a focus on statistical and neural-network learning algorithms that train on (annotated) text corpora to automatically acquire the knowledge needed to perform the task. Class lectures will discuss general issues as well as present abstract algorithms. The homework will touch both theoretical foundations of linguistic phenomena and implementation of the algorithms. Implemented versions of some of the algorithms will be provided in order to give a feel for how the systems discussed in class "really work" and allow for extensions and experimentation as part of the course projects.

Instructor: Nanyun (Violet) Peng (violetpeng@cs.ucla.edu)

Tentative Topics:
- Introduction to NLP: What's Important? What's Hard? What's Easy? Introduction to NLP applications, ambiguity in language, different levels of language.
- Lexical semantics: distributional semantics and word vectors, word-document matrix, LSA, neural network basics
- Language models: N-gram language model, log-linear language model, RNN language models, transformers, neural masked language model.
- Current NLP tasks and competitions: introductions to the current frontier of NLP applications, fairness considerations.

Lecture classroom: TBD

Undergrad Tutors:
- TBD
  - Cohort A
- TBD
  - Cohort B
- TBD
  - Cohort C
- TBD
  - Cohort D
- TBD
  - Cohort E
- TBD
  - Cohort F
- TBD
  - Cohort G
- TBD
  - Cohort H

Textbook (Optional):
1. Speech and Language Processing (3rd). Dan Jurafsky and James H. Martin
2. Foundations of Statistical Natural Language Processing, Chris Manning and Hinrich Schütze, MIT Press. Cambridge, MA.

Course material: Lectures, assignments and solutions will be posted on BruinLearn

Class Communication: Important class announcements will be done through online class forum on BruinLearn/Piazza. If you have any questions regarding class materials, they also need to be asked on Piazza.

Grading:
○   30% Homework
○   15% Course Project
○   20% Midterm
○   30% Final
○   5% Participation

You may discuss problems with friends, but you must write your solutions individually. I expect all students to follow the UCLA Student Conduct Code, which prohibits cheating, fabrication, and multiple submissions.

Tentative Schedule:

Week 1

	Morning (9:00 -- 11:50)	Afternoon (1:00 -- 4:00)	Deadlines
7/1	Introduction; what's NLP, why it is hard, what's generative AI	Introduction to Online Generative AI tools, Brainstorm Project Ideas	
7/2	Distributional Semantics and Word Embeddings	Project Discussion and Lab	Homework 1 out
7/3	N-gram Language Models	Project Discussion and Lab	Course Project Proposal Due
7/4	No Class	No Class	
7/5	Smoothing N-gram Language Models	Lab tour / seminar	

Week 2

	Morning (9:00 -- 11:50)	Afternoon (1:00 -- 4:00)	Deadlines
7/8	Intro to neural language models	Discussion and Lab	Homework 1 due Homework 2 out
7/9	Project Mid-Term Presentation	Project Mid-Term Presentation	
7/10	RNN language models	Discussion and Lab	Course Project: mid-term report due
7/11	Transformer Models	Discussion and Lab	
7/12	Masked Language Modeling	Lab tour / seminar	

Week 3

	Morning (9:00 -- 11:50)	Afternoon (1:00 -- 4:00)	Deadlines
7/15	Causal Language Models (i.e., ChatGPT)	Exam (90 mins); Discussion and Lab	Homework 2 due
7/16	Biases and Ethics Considerations for Generative AI	Discussion and Lab	

7/17	**Application: Creative Generation**	**Discussion and Lab**	
7/18	**Application: Multi-Modal Generation**	**Discussion and Lab**	**Course Project: final report due**
7/19	**Project Final Presentation + Demo**	**Course Project Presentation + Demo / Closing**	