IMPORTANT:
Please refer to the USC Center for Excellence in Teaching for current best practices in syllabus and course design. This document is intended to be a customizable template that primarily includes the technical elements required for the the purpose of central review by UCOC.

CSCI-699: Robustness and Generalization in NLP

Units: 4.0

Spring 2022, Monday/Wednesday, 4:00-5:50pm

IMPORTANT:

Location: Course website:
https://robinjia.github.io/classes/spring2022-csci699.html

Instructor: Robin Jia

Office: SAL 236

Office Hours: TBD

Contact Info: robinjia@usc.edu. I will reply within 48 hours. Please include "CSCI 699" in your email subject.

Course Description

In natural language processing (NLP), we set out to solve language-related *tasks* (e.g., machine translation, question answering) but often evaluate on narrow, in-distribution test *datasets*. With recent advances in deep learning, modern systems have achieved high accuracy on many canonical datasets, but still seem far from solving general tasks. In this class, we will survey recent research on robustness and generalization that studies this gap between in-distribution accuracy and task competency through out-of-distirbution settings. We will learn about different settings in which NLP systems often fail to generalize well, including adversarial perturbations, settings that require compositional reasoning, and domain transfer. We will also learn about how average accuracy can mask disparate performance across subpopulations, and how this can lead to undesirable consequences. Across these topics, we will cover methods both for measuring these robustness and generalization issues and ways that we can improve model robustness and generalization.

Learning Objectives

Students should come away with a broad understanding of the generalization challenges needed for NLP systems and how current research is attempting to solve these challenges. Taking this class will prepare students for research or other applied work in NLP and/or robustness. Students will get weekly practice analyzing and discussing current research papers.

Prerequisite(s): Familiarity with natural language processing and/or machine learning. Ideal pre/co-requisites are CSCI 544 (Applied Natural Language Processing) or CSCI 567 (Machine Learning). Email me if you want to enroll but are unsure if you meet the prerequisites.

Course Notes

Grading type: Letter of Credit/No Credit

Required Readings and Supplementary Materials

All required readings will be provided in PDF form. Optional recommended readings include:

- Lena Voita, "NLP Course For You." This is an excellent, relatively short introduction to modern NLP. I recommend starting here if you do not have prior NLP experience.
- Jurafsky and Martin, "Speech and Language Processing." This is my main recommendation for a comprehensive NLP textbook. The new 3rd edition is the most up-to-date NLP textbook available.
- Eisenstein. "Natural Language Processing." Some students may prefer this textbook. It focuses more on the machine learning and mathematical aspects of NLP.
- Barocas, Hardt, and Narayanan. "Fairness and Machine Learning: Limitations and Opportunities." For more about fairness and machine learning.

Description and Assessment of Assignments

Grades will be based on paper presentations (30%), discussion (10%), and a final project (60% total).

Paper presentations (30%)

Students will be expected to present 1-2 research papers and lead class discussion on these papers. The presentation should help everyone in the class understand these papers as well as relevant background material. The presenter should also prepare a few discussion questions to encourage discussion after the presentation.

Paper discussion participation (10%)

Students are expected to participate in class discussions. This includes asking questions during presentations as well as voicing opinions on discussion topics.

Final project (60% total)

Students must individually complete a final research project on a topic related to the class. This project is expected to include novel research on either (1) evaluation methodology for identifying problems with models related to robustness, generalization, or fairness, or (2) modeling innovations for improving robustness, generalization, fairness, or other related aspects of model behavior. Please come to office hours or email me if you have questions related to choosing a project direction.

All written assignments related to the final project should use the standard *ACL paper submission template.

Project proposal (5%). Students should submit a ~2-page proposal for their project by the end of Week 5. The proposal should describe the goal of the project and include a survey of related work.

Project progress report (10%). Students should submit a ~5-page progress report for their project by the end of Week 10. This should once again describe the project's goals (which may have changed since the proposal), initial results, and a concrete plan of what will be done for the final report. While the initial results need not be positive, students are expected to have made non-trivial implementation progress by this point.

Project final presentation (20%). This will be a 20-30 minute presentation during the last two weeks of class. Students should describe the motivation for their work, relevant background material, and results. I encourage students to present both positive and negative results. There will also be some time for audience questions.

Project final report (25%). Students should submit a ~8-page final report detailing all aspects of their project. The report should be structured like a conference paper, including an abstract, introduction, related work, and experiments. Parts of the proposal and progress report may be reused for the final report. Negative results will not be penalized, but should be accompanied with detailed analysis of why the proposed method did not work as anticipated.

Grading Breakdown

Assessment Tool (assignments)	% of Grade
Paper Presentations	30
Class Participation	10
Project proposal	5
Project progress report	10
Project final presentation	20
Project final report	25
TOTAL	100

Grading Scale

Course final grades will be determined using the following scale

A	95-100
A-	90-94
B+	87-89
B	83-86
B-	80-82
C+	77-79
C	73-76
C-	70-72
D+	67-69
D	63-66
D-	60-62
F	59 and below

Assignment Submission Policy

Assignments should be submitted by email before 11:59pm on the due date.

Grading Timeline

Assignments will be graded within one week of submission

Additional Policies

You are given **4 late days** to use for the project proposal and progress report (no late days for the final report), to be used in integer amounts and distributed as you see fit. Additional late days will result in a deduction of 10% of the grade on the corresponding assignment per day.

Course Schedule: A Weekly Breakdown

	Topics/Daily Activities	Readings/Preparation	Deliverables
Week 1	Introduction; The Turing Test		
Week 2	Adversarial examples		
Week 3	Adversarial perturbations, adversarial triggers		
Week 4	Model stealing, data poisoning; Introduction to domain adaptation		
Week 5	Domain-adaptive pretraining, fair generalization, empirical trends		**Project proposal due by Feb. 11 at 11:59pm PST**
Week 6	Spurious correlations, dataset biases		
Week 7	Avoiding spurious correlations at training time		
Week 8	Counterfactual data augmentation; Fairness in ML		
Week 9	Bias in NLP models		
Week 10	Distributionally robust optimization, bias amplification		**Project progress report due by March 25 at 11:59pm PST**
Week 11	Compositionality and systematicity		
Week 12	Improving compositional generalization; Adversarial data collection		
Week 13	Adversarial filtering; Conclusion		
Week 14	Project presentations		
Week 15	Project presentations		
FINAL	N/A		**Project final report due by May 6 at 11:59pm PST**

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 Research and Scholarship Misconduct.

Students and Disability Accommodations:

USC welcomes students with disabilities into all of the University's educational programs. The Office of Student Accessibility Services (OSAS) is responsible for the determination of appropriate accommodations for students who encounter disability-related barriers. Once a student has completed the OSAS process (registration, initial appointment, and submitted documentation) and accommodations are determined to be reasonable and appropriate, a Letter of Accommodation (LOA) will be available to generate for each course. The LOA must be given to each course instructor by the student and followed up with a discussion. This should be done as early in the semester as possible as accommodations are not retroactive. More information can be found at osas.usc.edu. You may contact OSAS at (213) 740-0776 or via email at osasfrontdesk@usc.edu.

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 for Equity, Equal Opportunity, and Title IX (EEO-TIX) - (213) 740-5086
eeotix.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 for Equity, Equal Opportunity, and Title for appropriate investigation, supportive measures, and response.

The Office of Student Accessibility Services (OSAS) - (213) 740-0776
osas.usc.edu
OSAS ensures equal access for students with disabilities through providing academic accommodations and auxiliary aids in accordance with federal laws and university policy.

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, Equity and Inclusion - (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.

*Occupational Therapy Faculty Practice - (323) 442-33*40 or otfp@med.usc.edu
chan.usc.edu/otfp
Confidential Lifestyle Redesign services for USC students to support health promoting habits and routines that enhance quality of life and academic performance.