
CS 333: NLP

Fall 2023

Prof. Carolyn Anderson
Wellesley College

PLEASE JOIN US FOR A



CS Holiday Party

ALL CS, MAS, DS AND CLCS
MAJORS AND MINORS.

THERE WILL BE TEA, SNACKS,
GAMES AND GINGERBREAD
HOUSES!

THURSDAY, DECEMBER 14, 2023

TIME: 4-6 PM

SCIENCE CENTER LEAKY BEAKER

[RSVP Here](#)



Disability Accommodations:
disabilityservices@wellesley.edu
or 781.283.2434



Questions:
cdelcour@wellesley.edu

Help hours

My help hours for the rest of the semester:

- ◆ Today: 3:30-4:30
- ◆ Monday: 3:30-4:30
- ◆ Thursday: 3-4

Final Project Presentations

- ◆ 2-3 minutes per person
- ◆ 1 slide (upload to Google Drive by 7am on Tuesday)

What Have We Learned?

Tasks We Learned This Semester

- ◆ Tokenization
- ◆ Sentence Segmentation
- ◆ Textual Similarity Analysis
- ◆ Representation Learning
- ◆ Classification: sentiment analysis, genre prediction...
- ◆ Language Modeling
- ◆ Common-sense reasoning

Techniques We Learned This Semester

- ◆ **Tokenization:** BPE
- ◆ **Textual Similarity Analysis:** tf-idf, PPMI
- ◆ **Representation Learning:** term-document frequency matrixes, static word embeddings, contextualized word embeddings
- ◆ **Classification:** Naive Bayes, regression, neural networks
- ◆ **Language Modeling:** n-grams, neural networks (feedforward, recurrent, attention-based)
- ◆ **Transfer Learning:** finetuning, prompt engineering

Models We Used This Semester

- ◆ word2vec
- ◆ DistilBERT
- ◆ feedforward neural networks in PyTorch
- ◆ LLaMA

Skills We Developed This Semester

- ♦ **text processing**: working with strings, reading in text datasets, tokenizing, cleaning text, regular expressions
- ♦ **evaluation**: computing metrics (accuracy, precision, recall, F-score, perplexity, and cosine similarity), understanding statistical reliability
- ♦ **implementing classification models**
- ♦ **building** and **running** neural network models
- ♦ **finetuning** and **prompting** large language models

What Is There Still
To Learn?

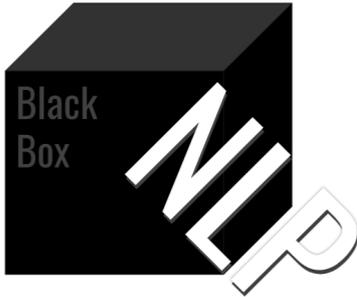
Interpretability and Explainability

How do blackbox models make decisions?

What are large language models learning, and how do they use this information?

What are the limits of what large language models?

How can we intervene when models are incorrect?



Black
Box

NLP

Workshop on analyzing and interpreting
neural networks for NLP

BlackboxNLP 2023

The sixth edition of BlackboxNLP will be co-located in Singapore on December 7 2023.

News

- We welcome EMNLP Findings papers with interpretability for presentation at the BlackboxNLP contact us via blackboxnlp@googlegroup.com October, noon ET.
- Find us on Twitter/X here: <https://twitter.com/blackboxnlp>
- We started a YouTube channel: <https://www.youtube.com/channel/UC...> Subscribe to be informed. You can already watch the BlackboxNLP

TrustNLP: Third Workshop on Trustworthy Natural Language Processing

Colocated with the Annual Conference of the Association for Computational Linguistics (ACL 2023)

About

Recent advances in Natural Language Processing, and the emergence of pretrained Large Language Models (LLM) specifically, have made NLP systems omnipresent in

Grounding and Multi-Modality

How do we combine text-based models with other modalities, like vision and perceptual information?

How can we help models relate to the physical world?

How do we help models understand social interaction?

How much of a role does access to perceptual information play in learning?

SpLU-RoboNLP 2023

Third International Combined Workshop on
Spatial Language Understanding and Grounded Communication for Robotics

[Topics](#) [Schedule](#) [Invited Speakers](#) [Organizers](#) [Program Committee](#)

Aim and Scope

Leveraging the foundation built in the prior workshops SpLU-RoboNLP 2021, SpLU 2020, SpLU-RoboNLP 2019, SpLU 2018 and RoboNLP 2017, we propose Spatial Language Understanding and Grounded Communication for Robotics. Natural language communication with general purpose embodied robots has long been a challenge, and natural language interfaces have the potential to make robots more accessible to a wider range of users. Achieving this goal requires the continuous development of new technologies for linking language to perception and action in the physical world. This joint workshop is aimed at bringing together the perspectives of researchers in natural language processing, robotics, human-computer interaction, and cognitive science to explore the challenges and opportunities of these systems and with human users, simulated embodied environments, and multimodal natural language and spatial language understanding to forge collaboration

[Confirmed Speakers](#) [Call For Papers](#) [Submission Instructions](#) [Schedule](#) [Papers](#) [Organizers](#)

NILLI

Novel Ideas in Learning-to-Learn through Interaction

All Editions

- [NILLI First Edition @ EMNLP'21](#)
- [NILLI Second Edition @ EMNLP'22](#)
- [NILLI Second Edition @ EMNLP'23](#)

Workshop @ EMNLP 2023

Collaborative dialogues [1, 2, 3] with automated systems through language interactions have become ubiquitous, wherein it is becoming common from setting an alarm to planning one's day through language

Linguistic Diversity

How can we improve NLP tools for low-resource languages?

How can we achieve good performance with sparse data?
With small models and limited compute?

How do models deal with different language varieties?



AmericasNLP Workshop
AmericasNLP 2024 Shared Task
AmericasNLP 2023 Workshop
AmericasNLP 2023 Shared Task
2022 Co

AmericasNLP

AmericasNLP aims to...

- ...encourage research on NLP, computational linguistics speech around the globe to work on native American languages
- ...connect researchers and professionals from underrepresented native speakers of endangered languages with the machine learning community

[AmericasNLP 2024 workshop](#)
AAACL 2024 in Mexico

[AmericasNLP 2023 workshop](#)
ACL 2023 in Toronto!
Results of the AmericasNLP Shared Task are now online

[AmericasNLP 2022 workshop](#)
ACL 2022 in Washington DC



COMPUTEL WORKSHOP

The Use of Computational Methods in the Study of Endangered Languages

[ComputEL Workshop](#) [ComputEL-7 Workshop \(2024\)](#) [Previous Workshops](#)



SustainNLP 2023

Fourth Workshop on Simple and Efficient Natural Language Processing

[Workshop Description](#)



Masakhane

A grassroots NLP community for Africa, by Africans

Our Mission

[Tweets by Masakhane](#)

Masakhane is a grassroots organisation whose mission is to strengthen and spur NLP research in African languages, for Africans, by Africans. Despite the fact that 2000 of the world's languages are African, African languages are barely represented in technology. The tragic past of colonialism has been devastating for African languages in terms of their support, preservation and integration. This has resulted in technological space that does not understand our names, our cultures, our places, our people.

Computational Social Science

How can we share our best approaches for text processing with other fields?

How well do NLP techniques work with limited data? In niche domains? In tasks that require expert judgment?



LaTeCH-CLfL 2023

The 7th Joint SIGHUM Workshop on Computational Linguistics, Cultural Heritage, Social Sciences, Humanities and Literature

to be held on 5 May 2023 in conjunction with EACL 2023 in Dubrovnik



[Link: Full program \(pdf\)](#)

Conference: November 9-10, 2023 at University of Massachusetts Amherst

The New Directions in Analyzing Text as Data (TADA) meeting is a leading forum for interdisciplinary research on the study of politics, society, and culture through computational analysis of documents. Recent advances in NLP have the potential to revolutionize how we study human society. But using these tools effectively, reliably, and equitably requires continuous dialog between experts across computational methods, social sciences, and the humanities.

[Registration](#)