NLP & SP Labs

Around the World

SUT
Speech Processing Lab

August
2021

NLP Groups Around the World

Some well-known labs are:

- The Stanford Natural Language Processing Group
- MIT NLP Group
- Cornell NLP group
- NLP @ Illinois
- Language Technologies Institute – Carnegie Mellon University
- Princeton NLP
- Natural Language Processing Group – Microsoft Research
- Computational Linguistics, University of Toronto
- Oxford Computational Linguistics Group
- Computer Laboratory, University of Cambridge
- Turing Center at University of Washington

The Stanford NLP Group

The Stanford NLP Group

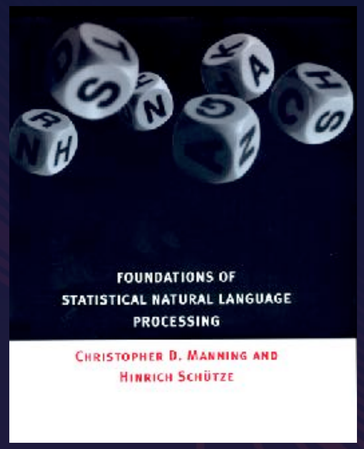
People

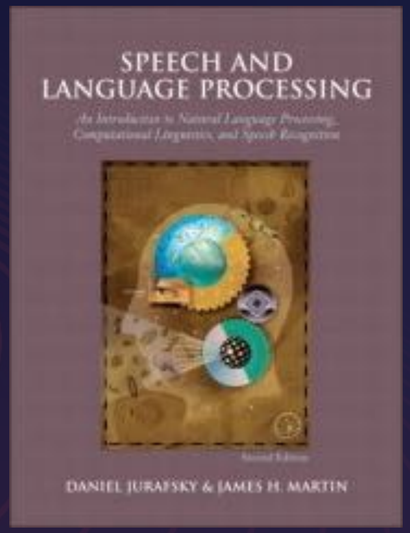
Publications

Research Blog

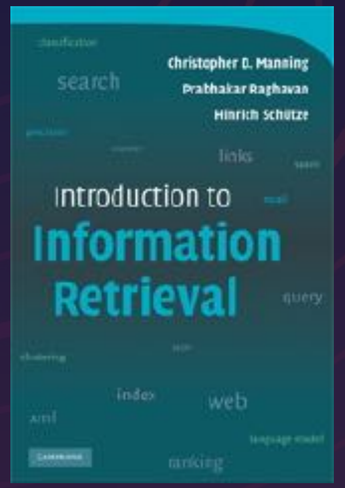
NLP Seminar

Dan Jurafsky





Chris Manning



The Stanford NLP Group

People

Publications

Research Blog

NLP Seminar

Postdocs

Antoine Bosselut
Computer Science

Dallas Card
Computer Science

Esin Durmus
Computer Science

Vivek Kulkarni
Computer Science

Alumni

Research Scientists

Michel Galley, *Computer Science*	Microsoft Research
Ramesh Nallapati, *Computer Science*	IBM Research
Mihai Surdeanu, *Computer Science*	Associate Professor, Schoo
	University of Arizona

Ph.D. Students

Gabor Angeli, *Computer Science*	Eloquent.ai
Sam Bowman, *Linguistics*	Assistant Professor in Lingu
Jason Brenier, *Linguistics*	Georgian Partners

The Stanford NLP Group

People

Publications

Research Blog

NLP Seminar

Hancheng Cao

How to pronounce my name?
In the Wade-Giles system of romanization, it is rendered as Hancheng Tsao.
In Chinese characters, it is 曹瀚成.

Follow @CaoHancheng
My LinkedIn Homepage

About Me

Hi! I am a third-year PhD candidate in Department of Computer Science, Stanford University (with a PhD minor in management science and engineering), advised by Prof. Dan McFarland and Prof. Dan Jurafsky. In previous quarters, I rotated with Prof. Michael Bernstein and Prof. Jure Leskovec, and collaborated with Prof. Melissa Valentine and Prof. Julien Clement. I am affiliated with Stanford NLP group and Stanford HCI group.

I received my bachelor degree (with honors) from Department of Electronic Engineering, Tsinghua University in 2018. I studied as an exchange student in A. James Clark School of Engineering, University of Maryland, College Park in Fall 2016.

I have been a research assistant in Future Communications and Internet Lab, Tsinghua University, advised by Prof. Yong Li and Prof. Vassilis Kostakos (University of Melbourne) since 2015. In Fall 2016, I was fortunate enough to work under the supervision of Distinguished University Professor Prof. Hanan Samet at University of Maryland. In Summer 2017, I was a visiting student and research assistant at Human Dynamics Group, MIT Media Lab, advised by Prof. Alex 'Sandy' Pentland and Prof. Xiaowen Dong.

Publications

- **H. Cao**, C. Lee, S. Iqbal, M. Czerwinski, P. Wong, S. Rintel, B. Hecht, J. Teevan, L. Yang, Large Scale Analysis of Multitasking Behavior During Remote Meetings, *in CHI 2021*. **Honorable Mention Award** [pdf][doi] [presentation] [Wired][Forbes][Microsoft Research Blog][LifeWire] [TechRepublic][Terra][NYTeknik]
- Z. Chen*, **H. Cao***, Y. Deng, X. Gao, J. Piao, F. Xu, Y. Zhang, Y. Li, Learning from Home: A Mixed-Methods Analysis of Live Streaming Based Remote Education Experience in Chinese Colleges during the COVID-19 Pandemic, *in CHI 2021*. [pdf][doi][presentation]

Education

Sep. 2018 - Jun. 2023 (Expected), Department of Computer Science, ***Stanford University***,
Doctor of Philosophy. Major in Computer Science, Minor in Management Science & Engineering.

Aug. 2014 - Jul. 2018, Department of Electronic Engineering, ***Tsinghua University***,
Bachelor of Engineering (with honors).

Aug. 2016 - Dec. 2016, A. James Clark School of Engineering, ***University of Maryland, College Park***,
Exchange Student.

Jun. 2017 - Sep. 2017, Human Dynamics Group, Media Laboratory, ***Massachusetts Institute of Technology***,
Visiting Student and Research Assistant.

Industrial Experience

Microsoft

Jun. 2021 - Sep. 2021, computational social science group, ***Microsoft Research***
Research Intern.
Hosts: Jake Hofman, Dan Goldstein.

The Stanford NLP Group

People

Publications

Research Blog

NLP Seminar

Title

Author(s)

Venue

1999 2016

Papers Books

Shikhar Murty, Tatsunori B Hashimoto and Christopher D Manning. 2021.
DReCa: A General Task Augmentation Strategy for Few-Shot Natural Language Inference.
Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL).
[paper, bib]

Douwe Kiela, Max Bartolo, Yixin Nie, Divyansh Kaushik, Atticus Geiger, Zhengxuan Wu, Bertie Vidgen, Grusha Prasad, Amanpreet Singh, Pratik Ringshia, Zhiyi Ma, Tristan Thrush, Sebastian Riedel, Zeerak Waseem, Pontus Stenetorp, Robin Mohit Bansal, Christopher Potts and Adina Williams. 2021.
Dynabench: Rethinking Benchmarking in NLP.
Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL).
[paper, bib]

Dorottya Demszky, Devyani Sharma, Jonathan H Clark, Vinodkumar Prabhakaran and Jacob Eisenstein. 2021.
Learning to Recognize Dialect Features.
Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL).
[paper, bib]

Yasuhide Miura, Yuhao Zhang, Emily Tsai, Curtis Langlotz and Dan Jurafsky. 2021.
Improving Factual Completeness and Consistency of Image-to-Text Radiology Report Generation.
Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL).
[paper, bib]

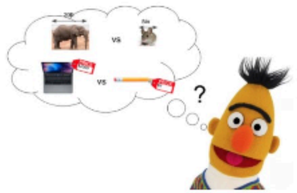
The Stanford NLP Group

People

Publications

Research Blog

NLP Seminar

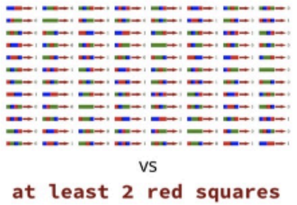


Do Language Models Know How Heavy an Elephant Is?

Xikun Zhang

Humans have a pretty good sense of scale, like the weight or price of different objects, but do pre-trained language representations? We measure how well pre-trained text representations capture scale and also come up with a improved version of BERT that captures scale better.

Continue reading



Learning from Language Explanations

Jesse Mu and Shikhar Murty

Language is a crucial way for humans to teach other humans. Can we use language to teach machines?

Continue reading

The Stanford NLP Group

People

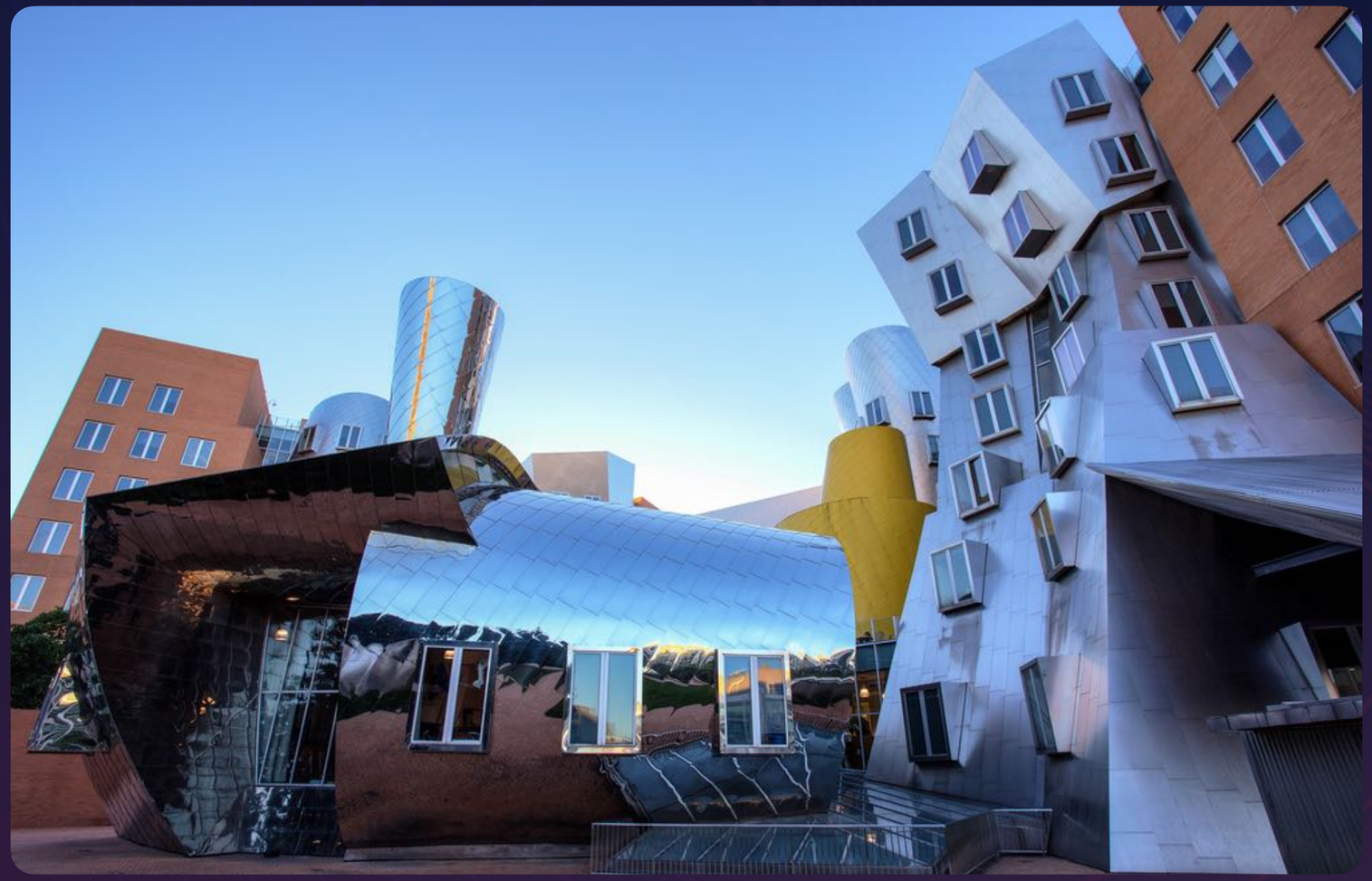
Publications

Research Blog

NLP Seminar

Date	Speaker	Title
Apr 1	Dzmitry Bahdanau	On Question Answering on Images and Databases (details)
Apr 8	Y-Lan Boureau	Better-behaved conversational agents (details)
Apr 15	Hinrich Schütze	Humans Learn From Task Descriptions and So Should Our Models (details, slides)
Apr 22	Tim Rocktäschel	Knowledge Intensive Reinforcement Learning (details, slides)
Apr 29	Anjalie Field	Detection of Stereotypes, Bias, and Prejudice in Text (details, slides)
May 6	Yacine Jernite	Making a Very Large Pretraining Dataset: some Social and Technical Considerations (details, slides)
May 13	Angelina McMillan-Major	Documenting Stochastic Parrots with Data Statements for NLP (details)
May 20	Martin Schrimpf	The neural architecture of language: Integrative modeling converges on predictive processing (details, slides)
May 27	Kalesha Bullard	Multi-Agent Reinforcement Learning towards Zero-Shot Emergent Communication (details)
Jun 3	Lillian Lee	On online discussion dynamics (details, slides)

NLP Group at MIT CSAIL



http://nlp.csail.mit.edu/

https://www.csail.mit.edu/research/natural-language-processing-group

NLP Group at MIT CSAIL

Publications

Events

The Limitations of Stylometry for Detecting Machine-Generated Fake News
Tal Schuster, Roei Schuster, Darsh J Shah, Regina Barzilay
Computational Linguistics journal, 2020.
[PDF] [Code]

Few-shot Text Classification with Distributional Signatures
Yujia Bao, Menghua Wu, Shiyu Chang, Regina Barzilay
International Conference on Learning Representations, 2020.
[PDF] [Code]

Automatic Fact-guided Sentence Modification
Darsh J Shah, Tal Schuster, Regina Barzilay
AAAI Conference on Artificial Intelligence (AAAI), 2020.
[PDF] [Code]

Capturing Greater Context for Question Generation
Luu Anh Tuan, Darsh J Shah, Regina Barzilay
AAAI Conference on Artificial Intelligence (AAAI), 2020.
[PDF]

NLP Group at MIT CSAIL

Publications

Events

Towards Debiasing Fact Verification Models
Tal Schuster, Darsh J Shah, Yun Jie Serene Yeo, Daniel Filizzola, Enrico Santus, Regina Barzilay
Empirical Methods in Natural Language Processing (EMNLP), 2019.
[PDF] [Code]

Working Hard or Hardly Working: Challenges of Integrating Typology into Neural Dependency Parsers
Adam Fisch, Jiang Guo, Regina Barzilay
Empirical Methods in Natural Language Processing (EMNLP), 2019.
[PDF] [Code]

Path-Augmented Graph Transformer Network
Benson Chen, Regina Barzilay, Tommi Jaakkola
ICML Workshop on Learning and Reasoning with Graph-Structured Representations, 2019.
[PDF] [Code]

NLP Group at MIT CSAIL

Publications

Events

Cross-Lingual Alignment of Contextual Word Embeddings, with Applications to Zero-shot Dependency Parsing
Tal Schuster, Ori Ram, Regina Barzilay and Amir Globerson
Proceedings of NAACL, 2019.
[PDF] [Code] [Video]

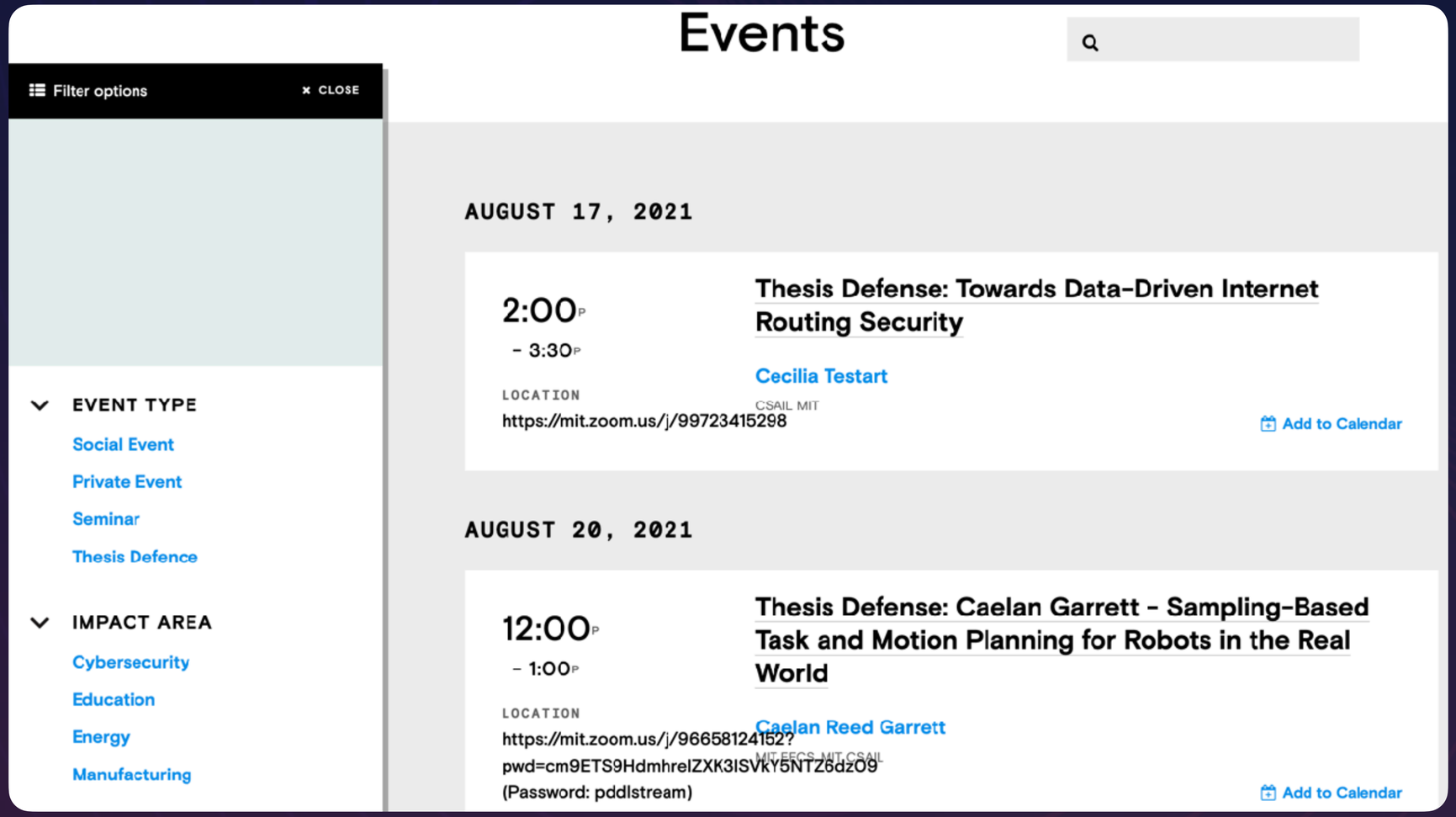
GraphIE: A Graph-Based Framework for Information Extraction
Yujie Qian, Enrico Santus, Zhijing Jin, Jiang Guo and Regina Barzilay
Proceedings of NAACL, 2019.
[PDF] [Code]

Inferring Which Medical Treatments Work from Reports of Clinical Trials
Eric Lehman, Jay DeYoung, Regina Barzilay and Byron C. Wallace
Proceedings of NAACL, 2019.
[PDF] [Code]

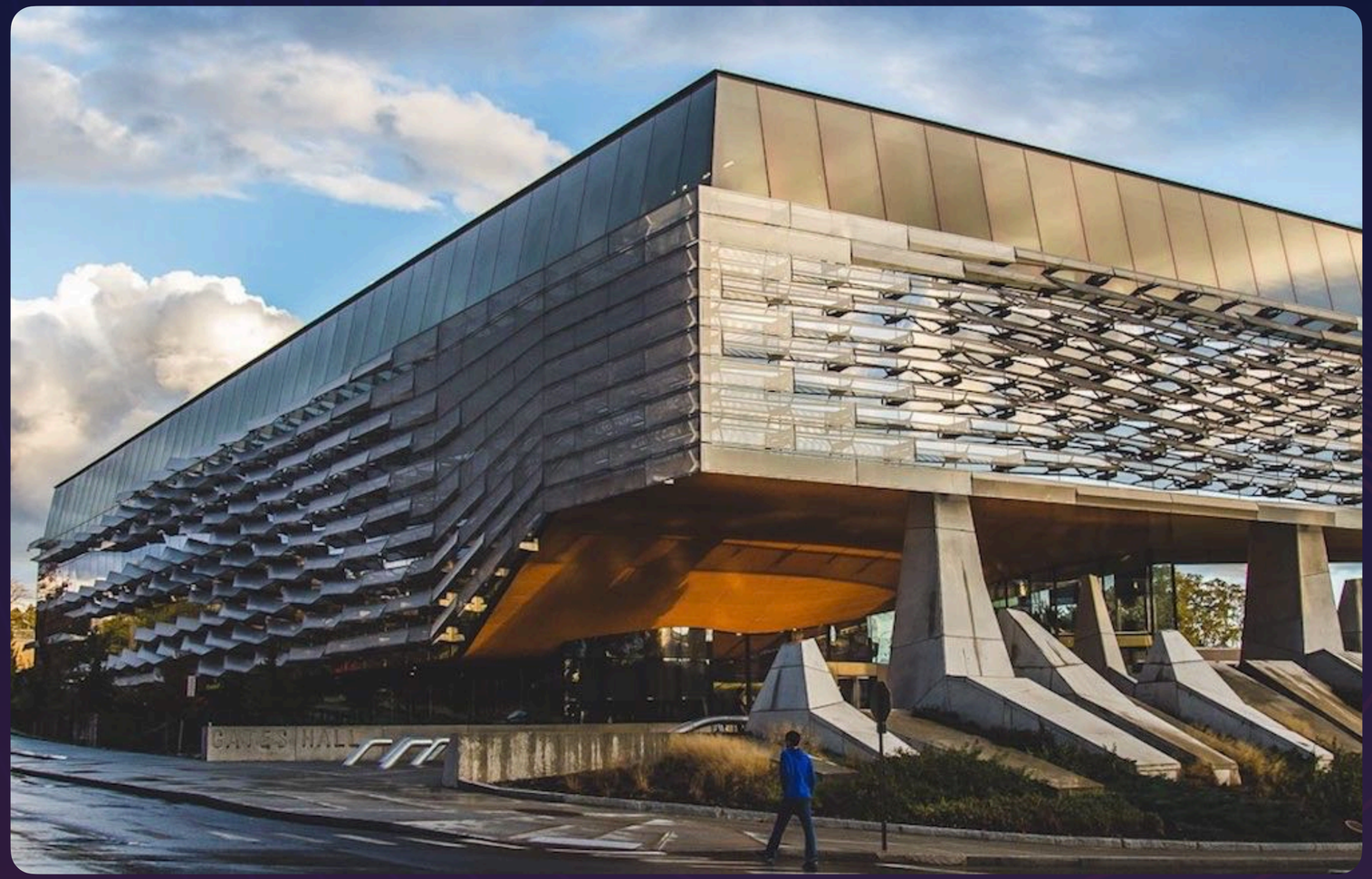
NLP Group at MIT CSAIL

Publications

Events



Cornell NLP group



Cornell NLP group

Publications

Courses

Computational Linguistics Lab

Resources

Epistemic Semantics in Guarded String Models
Eric Campbell and Mats Rooth.
2021.

Uncovering Constraint-Based Behavior in Neural Models via Targeted Fine-Tuning
Forrest Davis and Marten van Schijndel.
In *ArXiv*, 2021.

Template Filling with Generative Transformers
Xinya Du, Alexander M. Rush, and Claire Cardie.
In *NAACL*, 2021.

How many data points is a prompt worth?
Teven Le Scao and Alexander M. Rush.
In *NAACL*, 2021.

Adding Chit-Chat to Enhance Task-Oriented Dialogues
Kai Sun, Seungwhan Moon, Paul A. Crook, Stephen Roller Cardie.
In *NAACL*, 2021.

Cornell NLP group

Publications

Courses

Computational Linguistics Lab

Resources

Courses

Classes are listed in ascending numerical order: undergraduate courses appear before graduate courses.

INFO 3350 Text Mining History and Literature (also INFO 6350)

INFO 4300 Language and Information (also CS 4300)

LING 4424 Computational Linguistics (also CS 4744/COGST 4240)

PSYCH 4280 Computational Psycholinguistics (also COGST 4280/COGST 6280/LING 4428/LING 6628)

LING 4429 Grammar Formalisms (only offered 2016) (also LING 6429)

LING 4434 Computational Linguistics II

LING 4485 Topics in Computational Linguistics (also LING 6485)

Cornell NLP group

Publications

Courses

Computational Linguistics Lab

Resources

Projects

Neurolinguistics & Psycholinguistics

Using various different models from computational linguistics, we study the cognitive neuroscience of language. Through naturalistic speech comprehension data from fMRI studies, we are investigating different linguistic questions such as comparing compositional meaning to frozen meaning, semantic coherence vs. incoherence, binding theory & pronoun resolution among other topics.

Generalizable Learning

This project explores the learnability of various syntactic formalisms such as Categorial Grammars and Dependency Grammars and attempts to provide a more naturalistic algorithm for learning syntax that does not rely on extensively annotated structures, but rather on inferences that can be made from basic knowledge. We are also investigating how learned grammars compare to engineered grammars and to claims made in theoretical syntax.

Finite-state phonology

In this project we train a finite state model to detect prosodic cues in a speech corpus. We are specifically interested in detecting stress cues in Brazilian Portuguese and Bengali and finding empirical evidence for current theoretical views.

Cornell NLP group

Publications

Courses

Computational Linguistics Lab

Resources

More than 820 Language Corpora in 60+ languages (e.g. news text, dialogue corpora, television transcripts, etc)

Thank You for Your Attention!