

NLP4Science 2024

The 1st Workshop on NLP for Science

Proceedings of the Workshop

November 16, 2024

©2024 Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
317 Sidney Baker St. S
Suite 400 - 134
Kerrville, TX 78028
USA
Tel: +1-855-225-1962
acl@aclweb.org

ISBN 979-8-89176-185-8

Introduction

We are excited to welcome you to the NLP4Science workshop, part of EMNLP 2024, taking place in Miami, Florida, on November 16, 2024. This on-site workshop delves into an important frontier in natural language processing (NLP): leveraging NLP to deepen our understanding of the human mind and behavior.

In recent years, NLP and large language models (LLMs) have become pivotal tools in the scientific modeling of the human mind and behavior, influencing fields such as social science, psychology, psychiatry, healthcare, neuroscience, and behavioral economics. At NLP4Science, we will explore these advances through keynote talks, an expert panel, and a poster session featuring papers accepted to the workshop. Topics will include the principles of NLP-driven scientific modeling, methods for statistically robust evaluation of NLP models, experimental designs for NLP-based scientific research, and more.

This year, we received over 40 submissions, each carefully reviewed by our program committee. We are pleased to accept 22 papers, which will be presented as posters during the workshop. We extend our gratitude to all authors for their contributions, and we look forward to seeing how this work will advance our understanding of the human mind and behavior through NLP.

In addition to the poster session, we are honored to host several invited speakers who will present their cutting-edge research, including Amit Sharma (Microsoft Research), Rita Goldstein (Mount Sinai), Roger Levy (MIT), and Hadas Raviv (Princeton University). We will also hold a panel session where experts from multiple disciplines will discuss the future of NLP in scientific discovery. The panel, moderated by Roi Reichart (Technion), will feature Elliot Ash (ETH Zurich), Lyle Ungar (University of Pennsylvania), Diyi Yang (Stanford University), and Jekaterina Novikova (AI Risk and Vulnerability Alliance).

We would like to thank the organizing committee and workshop chairs, the authors for submitting their outstanding work, the reviewers for their hard work, and, of course, the workshop participants.

We hope this workshop fosters fruitful discussions and collaborations that push the boundaries of NLP's potential as a critical tool for scientific modeling of the human mind and behavior.

Organizing Committee

Organizing Committee

Lotem Peled-Cohen, Technion- Israel Institute for Technology
Nitay Calderon, Technion - Israel Institute for Technology
Alex Chapanin, Technion - Israel Institute for Technology
Rotem Dror, Haifa University
Amir Feder, Columbia University, Google Research
Ariel Goldstein, The Hebrew University
Anna Korhonen, University of Cambridge
Shir Lissak, Technion - Israel Institute for Technology
Yaakov Ophir, University of Cambridge
Ilanit Sobol, Technion - Israel Institute for Technology
Roi Reichart, Technion - Israel Institute for Technology
Refael Tikochinski, Technion - Israel Institute for Technology
Mor Ventura, Technion - Israel Institute for Technology

Program Committee

Program Committee

Dana Atzil-Slonim, Bar Ilan University, Israel
Aron Culotta, Tulane University, USA
Dan Goldwasser, Purdue University, USA
Justin Grimmer, Stanford University, USA
Vivek Gupta, University of Pennsylvania, USA
Nora Hollenstein, University of Copenhagen, Denmark
Qiao Jin MD, National Institutes of Health, USA
David Jurgens, Michigan University, USA
Jack LaViolette, Columbia University, USA
Marina Litvak, Shamoon College of Engineering, Israel
Emaad Manzoor, Cornell University, USA
Antonio Moreno Sandoval, Madrid University, Spain
Tristan Naumann, Microsoft
Reid Pryzant, Microsoft
Paul Rayson, Lancaster University, UK
Omer Shubi, Technion- Israel Institute for Technology
Hiroya Takamura, Tokyo Institute of Technology, Japan
Inbal Yahav, Tel Aviv University, Israel
Andrew Yates, University of Amsterdam, Netherlands
Zach Wood-Doughty, Northwest University, USA
Ayah Zirikly, Johns Hopkins University, USA
Carolina Zheng, Columbia University, USA

Keynote Talk

Teaching Causal Reasoning to Transformers

Amit Sharma
Microsoft Research, India



Sat, November 16, 2024 – Time: 09:00 – 09:45 –

Abstract: Large language models (LLMs) have demonstrated remarkable accuracy in identifying cause-and-effect relationships across various scientific fields. However, their ability to reason over these relationships remains inconsistent. In this work, we introduce an axiomatic approach to enhance causal reasoning in transformer models. By training on synthetic demonstrations of causal axioms, we show that even models with fewer than 100 million parameters can achieve reasoning capabilities comparable to much larger models, such as Phi and GPT-4. This suggests that axiomatic training offers a new method for imparting causal reasoning to LLMs, without the need for active interventions. Additionally, it may help explain how models like GPT-4, which are trained solely on observational data, can exhibit seemingly sophisticated reasoning.

Keynote Talk

Natural Language Processing of Spontaneous Speech in Drug Addiction: A Brain Biomarker of Treatment Effects and a Predictor of Outcomes

Rita Z. Goldstein
Icahn School of Medicine at Mount Sinai, USA



Sat, November 16, 2024 – Time: 09:45 – 10:30 –

Abstract: Escalating morbidity and mortality with opioid and stimulant use disorders makes it imperative to identify reliable and valid behavioral markers of brain function that change with treatment and predict outcomes. A naturalistic approach targeting spontaneous speech can serve as a readily accessible behavioral marker of neural plasticity with treatment, amenable for analyses with innovative artificial intelligence (AI), including natural language processing (NLP) approaches. Language, a ubiquitous, dense and measurable yet complex brain function that engages numerous higher-order executive functions (including the prefrontal cortex, PFC), is a distinctively human medium to express emotions, thoughts, and perceptions and to communicate. Concomitant with evidence for PFC recovery with treatment and abstinence in people with heroin and cocaine use disorders, using NLP our results show the impact of 3-months of treatment also on verbal movie-recall, further tracking the brain’s shared representation of this drug-themed complex stimulus, and its change with time; in contrast to common lab-based tasks, movies establish a dynamic narrative context, approximating real world processing by allowing a glimpse into the way groups of individuals with common addiction related experiences collectively attribute salience in their environment, addressing the social nature of drug use and recovery. Our results also show that NLP of speech about the consequences of drug use or abstinence outperformed numerous other models (including those incorporating drug use itself) as a predictor of 1-year outcomes, suggesting that hidden predictors of longitudinal craving, withdrawal, abstinence and drug use are embedded in peoples’ verbal outputs at baseline. The machine-learning community could only unlock the power of AI by first painstakingly curating large-scale training datasets. Using rich data that can be collected at scale, here we demonstrate a naturalistic ecologically valid approach to identify prognostic bio-markers of vulnerability and resilience in drug addiction. These bio-markers could ultimately facilitate the identification and design of neuroscience-informed empirically-based individually-tailored intervention and prevention approaches, timely deployed to improve outcomes, optimize recovery and save lives in people with drug addiction.

Keynote Talk

Cognitive Science of Language in the Era of Large Language Models

Roger Levy

Department of Brain and Cognitive Sciences at MIT, USA



Sat, November 16, 2024 – Time: 13:00 – 13:45 –

Abstract: Large language models (LLMs) are the first human-created artifacts whose text processing and generation capabilities seem to approach our own. But the hardware they run on is vastly different

than ours, and the software implementing them probably is too. How, then, can we use LLMs to advance the science of language in the human mind? In this talk I give examples of three key ways LLMs can contribute to the cognitive science of language: they can help us place lower bounds on the learnability of linguistic generalizations; they can help us reverse-engineer human language processing mechanisms; and they can help us develop hypotheses for the interface between language and other cognitive mechanisms.

Keynote Talk

The First 1000 Days - Harnessing Real-World Data to Model the Complexity of Language Acquisition

Hadas Raviv

Princeton Neuroscience Institute, USA



Sat, November 16, 2024 – Time: 13:45 – 14:30 –

Abstract: Human motor, cognitive, and social skills undergo extraordinary development during life's first 1000 days (1kD). Babies are born helpless, equipped with minimal cognitive-motor skills, and must rely on their caregivers to provide for their needs. However, within a few years, they walk, climb, talk, and reason about the world. Despite the large body of research devoted to studying the first 1000 days, our understanding of the developmental processes that shape children's future outcomes is limited. One of the biggest impediments to modeling human development is the lack of dense measurements of children's lives during their first few years of life. Until recently, capturing, managing, and analyzing large longitudinal datasets containing the rich sensory input with which the child is immersed has been technologically challenging.

Keynote Talk

The Role of NLP Interpretability in Advancing the Human Mind and Social Science Research

Nitay Calderon

The Faculty of Data and Decision Sciences at the Technion, Israel



Sat, November 16, 2024 – Time: 16:15 – 16:45 –

Abstract: The role of language in modern science goes far beyond its use as a tool for communicating research, sharing knowledge, or exploring new ideas. Language is fundamental to understanding human behavior, thoughts, and social development, serving as a gateway to comprehending the human mind. Given this centrality, it is no surprise that recent advances in Natural Language Processing (NLP) and the rise of Large Language Models (LLMs) have led researchers to leverage NLP for modeling scientific phenomena, decoding complex patterns, and uncovering insights about humanity. In this talk, I will explore how NLP interpretability methods are instrumental for the NLP4Science approach. Specifically, interpretability enables us to understand the underlying mechanisms and patterns the NLP model identifies, facilitating deeper comprehension and advancing scientific discoveries. I will review key interpretability paradigms, compare their strengths and limitations, and showcase their application in diverse scientific disciplines. For example, probing techniques are commonly used in neuroscience by aligning NLP representations with brain signals, and causal NLP methods such as counterfactuals are used in political science. I will then present a comprehensive analysis of over two thousand interpretability papers, highlighting trends in NLP interpretability paradigms and comparing them to those in other fields. Finally, I will discuss the broader implications of these findings and offer recommendations for promoting the adoption of NLP interpretability techniques in fields beyond NLP.

Table of Contents

<i>TokenSHAP: Interpreting Large Language Models with Monte Carlo Shapley Value Estimation</i> Miriam Horovicz and Roni Goldshmidt	1
<i>Prediction of CRISPR On-Target Effects via Deep Learning</i> Condy Bao and Fuxiao Liu	9
<i>What an Elegant Bridge: Multilingual LLMs are Biased Similarly in Different Languages</i> Viktor Mihaylov and Aleksandar Shtedritski	16
<i>PsychoLex: Unveiling the Psychological Mind of Large Language Models</i> Mohammad Amin Abbasi, Farnaz Sadat Mirnezami and Hassan Naderi	24
<i>Two-Stage Graph-Augmented Summarization of Scientific Documents</i> Rezvaneh Rezapour, Yubin Ge, Kanyao Han, Ray Jeong and Jana Diesner	36
<i>GCD-TM: Graph-Driven Community Detection for Topic Modelling in Psychiatry Texts</i> Anusuya Krishnan and Isaias Mehari Ghebrehiwet	47
<i>SCITUNE: Aligning Large Language Models with Human-Curated Scientific Multimodal Instructions</i> Sameera Horawalavithana, Sai Munikoti, Ian Stewart, Henry Kvinge and Karl Pazdernik	58
<i>RACER: An LLM-powered Methodology for Scalable Analysis of Semi-structured Mental Health Interviews</i> Satpreet Harcharan Singh, Kevin Jiang, Kanchan Bhasin, Ashutosh Sabharwal, Nidal Moukaddam and Ankit Patel	73
<i>Soft Measures for Extracting Causal Collective Intelligence</i> Maryam Berijanian, Spencer Dork, Kuldeep Singh, Michael Riley Millikan, Ashlin Riggs, Aadarsh Swaminathan, Sarah L. Gibbs, Scott E. Friedman and Nathan Brugnone	99
<i>Hypothesis Generation with Large Language Models</i> Yangqiaoyu Zhou, Haokun Liu, Tejes Srivastava, Hongyuan Mei and Chenhao Tan	117
<i>Dreaming with ChatGPT: Unraveling the Challenges of LLMs Dream Generation</i> Harel Berger, Hadar King and Omer David	140
<i>LLMs and NLP for Generalized Learning in AI-Enhanced Educational Videos and Powering Curated Videos with Generative Intelligence</i> Naina Chaturvedi	148
<i>The Moral Foundations Weibo Corpus</i> Renjie Cao, Miaoyan Hu, Jiahao Wei and Baha Ihnaini	155
<i>Why So Serious: Humor and its Association with Treatment Measurements Process and Outcome</i> Matan Kenigsbuch and Natalie Shapira	166
<i>Learning the Bitter Lesson: Empirical Evidence from 20 Years of CVPR Proceedings</i> Mojtaba Yousefi and Jack Collins	175
<i>Personalized-ABA: Personalized Treatment Plan Generation for Applied Behavior Analysis using Natural Language Processing</i> Aman Kumar, Mareiko Au, Raj Semlawat, Malavica Sridhar and Hitesh Gurnani	188

<i>Exploring Scientific Hypothesis Generation with Mamba</i> Miaosen Chai, Emily Herron, Erick Cervantes and Tirthankar Ghosal	197
<i>Benchmarking Automated Theorem Proving with Large Language Models</i> Vanessa Lama, Catherine Ma and Tirthankar Ghosal	208
<i>The Grid: A semi-automated tool to support expert-driven modeling</i> Allegra A. Beal Cohen, Maria Alexeeva, Keith Alcock and Mihai Surdeanu	219
<i>Categorical Syllogisms Revisited: A Review of the Logical Reasoning Abilities of LLMs for Analyzing Categorical Syllogisms</i> Shi Zong and Jimmy Lin	230
<i>Individuation in Neural Models with and without Visual Grounding</i> Alexey Tikhonov, Lisa Bylinina and Ivan P. Yamshchikov	240
<i>CogErgLLM: Exploring Large Language Model Systems Design Perspective Using Cognitive Ergonomics</i> Azmine Toushik Wasi and Mst Rafia Islam	249

Program

Saturday, November 16, 2024

- 08:45 - 09:00 *Gathering and Welcome*
- 09:00 - 09:45 *Invited Speaker 1 - Amit Sharma*
- 09:45 - 10:30 *Invited Speaker 2 - Rita Goldstein*
- 10:30 - 11:00 *Coffee Break*
- 11:00 - 12:00 *Panel Discussion & Q&A*
- 12:00 - 13:00 *Lunch Break*
- 13:00 - 13:45 *Invited Speaker 3 - Roger Levy*
- 13:45 - 14:30 *Invited Speaker 4 - Hadas Raviv*
- 14:30 - 16:00 *Poster Session + Coffee Break*
- 16:00 - 16:15 *Best Paper Announcement + Short Oral*
- 16:15 - 16:45 *Invited Speaker 5 - Nitay Calderon*
- 16:45 - 17:00 *Closing Remarks*