

LNLS 2022

**The First Workshop on Learning with Natural Language  
Supervision**

**Proceedings of the Workshop**

May 26, 2022

The LNLS organizers gratefully acknowledge the support from the following sponsors.

**Gold**



©2022 Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)  
209 N. Eighth Street  
Stroudsburg, PA 18360  
USA  
Tel: +1-570-476-8006  
Fax: +1-570-476-0860  
[acl@aclweb.org](mailto:acl@aclweb.org)

ISBN 978-1-955917-45-2

## Introduction

To a growing extent, advances across machine learning application domains are driven by advances in NLP. In computer vision, image captions are used to shape learned representations of images [Frome et al., 2013, Mu et al., 2020, Radford et al., 2021, Desai and Johnson 2021]. In programming languages, textual code comments are used to guide and constrain models for example-based program synthesis [Yaghmazadeh et al., 2017, Austin et al., 2021, Wong et al., 2021]. In robotics and more general policy learning settings, rules and instructions are used to enable generalization to new environments and goals [Zhong et al., 2020, Narasimhan et al., 2018, Sharma et al., 2020]. Within NLP, rich natural-language annotations and task descriptions are used to improve the performance and interpretability of models for text categorization and question answering [Hancock et al., 2018, Weller et al., 2020, Efrat et al., 2020]. And in cognitive science, experimental evidence suggests that language shapes many other aspects of human cognition (e.g. Jones et al., 1991).

At present, however, most research on learning from language takes place within individual application domains (and mostly outside of the NLP community). While many approaches to language supervision are domain-general, and closely connected to “core” NLP research, there are currently no venues where researchers from across the field can meet to share ideas and draw connections between their disparate lines of research. Our workshop will offer a central meeting point for research on language-based supervision, enabling researchers within and beyond NLP to discuss how language processing models and algorithms can be brought to bear on problems beyond the textual realm (e.g. visual recognition, robotics, program synthesis, sequential decision making). Existing workshops like RoboNLP, SPLU, and ViGiL focus on models for multi-modality; inspired by the relationship between language and human cognitive development, our workshop will emphasize broader use of language not just as an input modality but a fundamental source of information about the structure of tasks and problem domains.

In keeping with this interdisciplinary focus, our workshop format differs in two ways from a standard NLP workshop: first, with a special emphasis on speakers and attendees who would not typically attend NLP conferences; second, by replacing the standard panel discussion with a series of workshop-wide breakout sessions aimed at seeding cross-institutional collaborations around new tasks, datasets, and models.

# Organizing Committee

## Organizing Committee

Jacob Andreas, MIT, USA

Karthik Narasimhan, Princeton University, USA

Aida Nematzadeh, DeepMind, UK

# Program Committee

## Program Chairs

Jacob Andreas, Massachusetts Institute of Technology and Microsoft  
Karthik R Narasimhan, Princeton University  
Aida Nematzadeh, Deepmind

## Reviewers

Siddharth Karamcheti, Stanford University  
Theodore Sumers, Princeton University  
Afra Feyza Akyürek, Boston University  
Bishan Yang, Carnegie Mellon University  
Jesse Mu, Stanford University  
Catherine Wong, Massachusetts Institute of Technology  
Erin Grant, University of California Berkeley  
Ekin Akyürek, Massachusetts Institute of Technology  
Pratyusha Sharma, Massachusetts Institute of Technology  
Ameet Deshpande, Princeton University  
Evan Hernandez, Massachusetts Institute of Technology  
Olivia Watkins, University of California, Berkeley  
Robert D. Hawkins, Princeton University  
Shunyu Yao, Princeton University  
Rakesh R Menon, Department of Computer Science, University of North Carolina, Chapel Hill

## Table of Contents

<i>Finding Sub-task Structure with Natural Language Instruction</i> Ryokan Ri, Yufang Hou, Radu Marinescu and Akihiro Kishimoto .....	1
<i>GrammarSHAP: An Efficient Model-Agnostic and Structure-Aware NLP Explainer</i> Edoardo Mosca, Defne Demirtürk, Luca Mülln, Fabio Raffagnato and Georg Groh .....	10
<i>Single-Turn Debate Does Not Help Humans Answer Hard Reading-Comprehension Questions</i> Alicia Parrish, Harsh Trivedi, Ethan Perez, Angelica Chen, Nikita Nangia, Jason Phang and Samuel R. Bowman .....	17
<i>When Can Models Learn From Explanations? A Formal Framework for Understanding the Roles of Explanation Data</i> Peter Hase and Mohit Bansal .....	29
<i>A survey on improving NLP models with human explanations</i> Mareike Hartmann and Daniel Sonntag .....	40