Carnegie Mellon University
Language Technologies Institute

Percy Liang

Stanford University

Percy Liang is an Associate Professor of Computer Science at Stanford University (B.S. from MIT, 2004; Ph.D. from UC Berkeley, 2011). His research spans machine learning and natural language processing, with the goal of developing trustworthy agents that can communicate effectively with people and improve over time through interaction. Specific topics include question answering, dialogue, program induction, interactive learning, and reliable machine learning. His awards include the IJCAI Computers and Thought Award (2016), an NSF CAREER Award (2016), a Sloan Research Fellowship (2015), and a Microsoft Research Faculty Fellowship (2014).

Taming Deep Models and on Shaping their Development

Models in deep learning are wild beasts: they devour raw data, are powerful but hard to control. This talk explores two approaches to taming them. First, I will introduce concept bottleneck networks, in which a deep neural network makes a prediction via interpretable, high-level concepts. We show that such models can obtain comparable accuracy with standard models, while offering the unique ability for a human to perform test-time interventions on the concepts. Second, I will introduce prefix-tuning, which allows one to harness the power of pre-trained language models (e.g., GPT-2) for text generation tasks. The key idea is to learn a continuous task-specific prefix that primes the language model for the task at hand. Prefix-tuning obtains comparable accuracy to fine-tuning, while only updating 0.1% of the parameters. Finally, I will end with a broad question: what kind of datasets should the community develop to drive innovation in modeling approaches? Are size and realism necessary attributes of a dataset? Could we have made all the modeling progress in NLP without SQuAD? As this counterfactual question is impossible to answer, we perform a retrospective study on 20 modeling approaches and show that even a small, synthetic dataset can track the progress that was made on SQuAD. While inconclusive, this result encourages us to think more critically about the value of datasets during their construction.

Friday, February 26, 2021
2:20 - 3:40 PM EST

Join the meeting on Zoom
Meeting ID 935 3287 1380 Passcode 546823

LTI Colloquium 2020-21
is generously sponsored by abridge