THE NANCY IDE ANNUAL LECTURE IN COMPUTER SCIENCE

USING LARGE LANGUAGE MODELS TO BUILD EXPLAINABLE CLASSIFIERS

April 5, 2024
5:30 PM
Rocky 300

Chris Callison-Burch

This presentation discusses research on using large language models (LLMs) to build explainable classifiers. It will show off work from my PhD students and collaborators on several recent research directions:

- Image classification with explainable features (arxiv.org/abs/2211.11158)
- Text classification with explainable features (arxiv.org/abs/2305.12696 and arxiv.org/abs/2310.19660)
- The importance of faithfulness in explanations (arxiv.org/abs/2209.11326)
- A faithful "chain of thought" LLM reasoner that produces code in its explanations (arxiv.org/abs/2301.13379)

The talk will cover joint work with: Marianna Apidianaki, Liam Dugan, Shreya Havaldar, Daniel Jin, Ansh Kothary, Veronica Qing Lyu, Kathleen McKeown, Josh Ludan, Artemis Panagopoulou, Ajay Patel, Delip Rao, Adam Stein, Eric Wong, Yue Yang, Mark Yatskar, Harry Li Zhang, Shenghao Zhou and others.

Chris Callison-Burch is an associate professor of Computer and Information Science at the University of Pennsylvania. His course on Artificial Intelligence has one of the highest enrollments at the university with over 500 students taking the class each Fall. He is best known for his research into natural language processing. His current research is focused on applications of large language models to long-standing challenges in artificial intelligence. His PhD students joke that now whenever they ask him anything his first response is “Have you tried GPT for that?” Prof Callison-Burch has more than 100 publications, which have been cited over 25,000 times. He is a Sloan Research Fellow, and he has received faculty research awards from Google, Microsoft, Amazon, Facebook, and Roblox, in addition to funding from DARPA, IARPA, and the NSF. In 2023, Prof Callison-Burch testified before congress about the relationship of generative AI and Copyright Law.