

## **PLI seed grant recipients present at symposium on their large language model research**

Allison Gasparini

On May 2, the Princeton Language and Intelligence (PLI) initiative held a spring symposium, highlighting the groundbreaking projects of [PLI Seed Grant recipients](#) who have incorporated large language models into their research.

At the initiative's [first symposium](#), held in November, Sanjeev Arora, Princeton's Charles C. Fitzmorris Professor in Computer Science and Director of PLI, emphasized, "The goal today is to showcase how researchers are using or plan to use large AI models to further their research." The PLI symposium is one way the initiative shows its dedication to promoting interdisciplinary collaboration and enhancing a broader community understanding of AI.

### **LLMs across disciplines**

For the symposium, seven faculty members presented 5-minute lightning talks on the ways LLMs are informing or aiding their research.

From the Department of Anthropology and the Princeton Writing Program, lecturer Andrea DiGiorgio is using machine learning methods to understand the impacts of social media on wildlife conservation efforts. Well-meaning wildlife research, conservation, and rescue groups might post cute videos of themselves interacting with endangered animals to social media channels, but a previous scientific study found that these types of images may inspire conservation-negative thinking – such as increasing a viewer's desire to own such an animal as a pet. "My research investigates the potential pitfalls of conservation social media in an attempt to understand how we can craft the most productive and positive marketing," said DiGiorgio.

Using a small language processing model, DiGiorgio and colleagues analyzed the comments on videos posted of rescued orangutans. While videos of baby orangutans and orangutans interacting with humans got the most views, they garnered more conservation-negative comments. This summer, DiGiorgio is planning to run wildlife social media posts and their corresponding comments through an AI model to further analyze the association between images and the comments they inspire. With her findings she hopes to "make recommendations to wildlife and conservation agencies and researchers about how they can create social media that doesn't inadvertently garner these counterproductive viewer responses."

Christine Fellbaum, lecturer with the rank of professor in the Council of the Humanities, the Program in Linguistics, Freshman Seminars, and Computer Science is looking to create better representation of African languages in LLM and Natural Language Processing research. Of the 7,000 languages spoken in the world, 2,000 are spoken in Africa. In comparison, African languages are far underrepresented in available pre-training datasets. "Our objective is to create a quality dataset with theoretically sound and consistent syntactic human annotations for 11 typologically diverse African languages," said Fellbaum.

Fellbaum's collaborator Happy Buzaaba, a postdoctoral researcher in the Center for Digital Humanities, further said the research team plans to annotate 1500 sentences per language for the dataset. "We anticipate this [dataset] to facilitate NLP research, language technology, artificial intelligence, and large language models for underrepresented languages," said Fellbaum.

Associate Professor of Politics and Public Affairs, Princeton School of Public and International Affairs Jonathan Mummolo said he's been looking at police body camera footage as a data source for understanding civilian-police interactions. "A broader theme in my work is focused on how to make valid statistical inferences about police-civilian interactions given that most of the data sources that we've used in the study of policing are incomplete or distorted in some way," said Mummolo.

For his research, Mummolo and colleagues are using a combination of human annotations and OpenAI system GPT-4V to develop an open source model to summarize the contents of this footage. "Until now, we have basically had to rely on self reports from officers to study these situations," Mummolo said. "If successful, this system would have widespread applications for both police oversight and academic research on police behavior."

Other talks during the symposium included from faculty in the Departments of Environmental and Civil Engineering, Molecular Biology, Operations Research and Financial Engineering, and more. The [presentations](#) are available to view as videos on the PLI website now.

Following the lightning talks, participants enjoyed a reception and a poster session which featured graduate students and postdoctoral researchers whose work involves the use of LLMs. PLI Seed Grants, awarded annually, are designed to promote the integration of large AI models in research activities. A call for 2024-25 proposals will be announced in the fall.