UNC Symposium on AI and Society

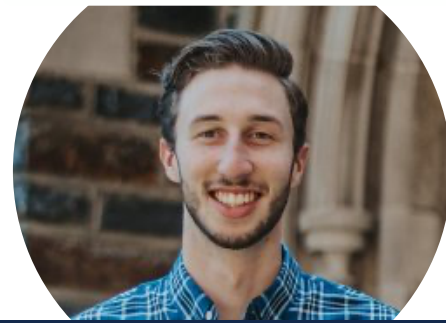**Algorithmic Monoculture and the Ethics of Systemic Exclusion**

Kathleen Creel
Northwestern

(When) Is It What's On The Inside That Counts? Balancing Internalist and Externalist Considerations in LLM Metasemantics

Josh Dever
UT- Austin

Toward Safe LLMs: Interpretability, Model Editing, and Scalable Oversight"

Peter Hase
UNC-CH

Addressing two classic debates in cognitive science with AI

Brenden Lake
NYU

Risk Assessment, Safety Alignment, and Guardrails for Generative Models

Bo Li
University of Chicago

Artificial Social Intelligence? On the challenges of Socially Aware and Ethically informed LLMs

Maarten Sap
Carnegie Mellon

New Models for Trustworthy AI, Norm Deviation, and Consent in Responsible Autonomy

Munindar P. Singh
NC State

Moral AI and How We Get There

Jana Schaich Borg &
Walter Sinnott-Armstrong
Duke

This event features discussions on the impact of AI on society, including topics such as ethical and moral AI, trustworthy and interpretable Large Language Models (LLMs), and social intelligence, led by experts across disciplines including computer science, philosophy, cognitive science, psychology, and data science. Join us for presentations of the latest work on AI and contribute to the dialogue on AI's role in society.

HOSTED BY: Mohit Bansal, *John R. & Louise S. Parker Professor* | Thomas Hofweber, *Professor*

UNC | COLLEGE OF ARTS AND SCIENCES
Computer Science

UNC | COLLEGE OF ARTS AND SCIENCES
Philosophy

April 25 & 26, 2024 | 141 Brooks Bldg.

WITH SUPPORT FROM: UNC | SCHOOL OF DATA SCIENCE AND SOCIETY