



Joint Institutskolloquium and Munich Econometrics Seminar

Causal Inference on Outcomes Learned from Text

Jann Spiess

(Stanford University)

10/06/2026, 4:15 pm

Department of Statistics, Ludwigstr. 33, Room 144

and online via Zoom ([Link](#))

(Meeting-ID: 631 1190 7291; Password: StatsCol)

We propose a machine-learning tool that yields causal inference on text in randomized trials. Based on a simple econometric framework in which text may capture outcomes of interest, our procedure addresses three questions: First, is the text affected by the treatment? Second, which outcomes is the effect on? And third, how complete is our description of causal effects? To answer all three questions, our approach uses large language models (LLMs) that suggest systematic differences across two groups of text documents and then provides valid inference based on costly validation. Specifically, we highlight the need for sample splitting to allow for statistical validation of LLM outputs, as well as the need for human labeling to validate substantive claims about how documents differ across groups. We illustrate the tool in a proof-of-concept application using abstracts of academic manuscripts.

About the Speaker:

Jann Spiess is an Associate Professor of Operations, Information & Technology at the Graduate School of Business, Stanford University. His current research focuses broadly on two related themes: (i) high-dimensional and robust causal inference, including work on using machine learning to improve inferences from randomized trials, robust inference in panel data, synthetic control, matching estimation, highly over-parametrized models, and high-dimensional outcome data; and (ii) data-driven decision-making with misaligned objectives, including work on algorithmic fairness, human-AI interaction, the regulation of algorithms, and the design of pre-analysis plans. Jann holds a PhD in economics from Harvard University. He studied mathematics at the University of Cambridge and the Technical University of Munich.

References:

Modarressi, I., Spiess, J., & Venugopal, A. (2025). Causal inference on outcomes learned from text. *arXiv preprint arXiv:2503.00725*. [[URL](#)]
