MARKETING COLLOQUIA

Sanjog Misra

Charles H. Kellstadt Distinguished Service Professor of Marketing and Applied AI

University of Chicago, Booth School of Business

FOUNDATION PRIORS

Foundation models (and in particular, Large language models) can generate informative responses, prompting growing interest in using these "synthetic" outputs as data in empirical research and decision-making. This paper introduces the idea of the foundation prior, which treats model-generated outputs not as real observations but rather as draws from a distribution that reflects both the model's learned patterns as well as the user's own priors, expectations, and biases. We model the subjectivity involved in the generative process and make explicit the dependence of the generative outputs on the user's anticipated data distribution, the prompt-engineering process, and the trust placed in the foundation model. We characterize the structure of the foundation prior via a generalized Bayesian framework and outline how such synthesized data can be incorporated into traditional statistical and analytical frameworks. We follow up with a number of use cases and applications that offer practical ways to use synthetic data responsibly, whether it be to refine complex models, augment latent constructs, guide experimental design, or support robust decision-making.

Wharton UNIVERSITY of PENNSYLVANIA

MARKETING DEPARTMENT