Yohei Nishimura

Ph.D. Candidate, Department of Marketing

University of Wisconsin-Madison

Email: ynishimura@wisc.edu | **Phone:** +1-608-421-4995

Website: Google Scholar ◇ GitHub ◇ LinkedIn ◇ Personal Website

Education

University of Wisconsin-Madison — Madison
Ph.D. Candidate in Quantitative Marketing — Sep, 2023 – Present
(Expected Graduation: May, 2028)

University of Wisconsin-Madison — Madison
M.S. in Computer Sciences — Sep, 2021 – May, 2023

University of Tokyo — Tokyo
B.A. in Economics — Apr, 2003 – Mar, 2007

Graduate Study Profile

Current Status:
Ph.D. Candidate in Quantitative Marketing, Wisconsin School of Business
Enrollment Year: Currently in 3rd year (Enrolled in Fall 2023)

Coursework Description:
I have successfully completed all core Ph.D. coursework, including Microeconomics and Econometrics sequences. In addition, I have completed courses in:

- Industrial Organization
- Applied Econometrics including Causal Inference
- Machine Learning, Deep Learning on Vision and Text Data
- Theory and Application of Reinforcement Learning, Dynamic Optimization
- Semiparametric Theory and Methods
- Bayesian Statistics

Research Interests

My primary research interests lie at the intersection of **Digital Economics**, **Quantitative Marketing**, and **Artificial Intelligence**. I combine structural econometric models with machine learning techniques, specifically Large Language Models (LLMs) and deep generative models, to study:

- **The Creator Economy & Platform Design:** Two-sided matching markets, influencer marketing, and algorithmic allocation efficiency.
- **GenAI in Business Processes:** Human-AI collaboration, the impact of LLMs on research productivity, and automated content generation.

- **Methodological Innovation:** Integrating unstructured data (text/image embeddings) into structural estimation and using generative AI to solve data scarcity problems (e.g., informative priors).

Research Projects

Prospective Job Market Paper:

From Match to Impact: Driving Ad Effectiveness in Micro Influencer Marketing

Abstract:

In the influencer marketing industry, advertisers often select influencers based on observable metrics like follower counts, potentially overlooking the "fit" between the influencer's content and the brand's positioning. This paper investigates this misallocation by developing and estimating a two-sided structural matching model.

I model influencers' application and advertisers' offer utilities as functions of campaign rewards, observable attributes, and match-specific synergy terms. A key methodological innovation is the construction of these synergy terms: I convert high-dimensional textual and visual content into low-dimensional representations via deep-learning embeddings and map them into economically interpretable match variables. The model links endogenous match formation to realized engagement outcomes, while addressing selection bias arising from the two-sided matching process through a correlated error structure.

Counterfactual simulations indicate that current platform mechanisms generate systematic misallocation. I show that implementing Gale-Shapley type centralized matching algorithms can materially increase aggregate engagement and surplus for both sides. This work demonstrates how structural models can integrate unstructured data to evaluate platform interventions in digital advertising markets.

Published & Revised Work:

1. AI-Human Hybrids for Marketing Research: Leveraging LLMs as Collaborators

(with N. Arora and I. Chakraborty)

Journal of Marketing** (Vol 89, 2025) – **Published, Citation: 107 on Nov. 23

We propose that a human–LLM hybrid approach leads to significant efficiency and effectiveness gains in the marketing research process. In qualitative research, we show that LLMs can effectively create sample characteristics, generate synthetic respondents, and moderate in-depth interviews. Our results demonstrate that the AI–human hybrid generates information-rich, coherent data that surpasses human-only data in depth and insightfulness.

In quantitative domains, we find that the quality of synthetic data improves significantly through few-shot learning and retrieval-augmented generation (RAG). By collaborating with a Fortune 500 food company to replicate a prior study using GPT-4, we provide experimental evidence that humans and LLMs possess complementary skills, and that the hybrid model consistently outperforms human-only or LLM-only counterparts.

2. Leveraging Generative AI to Create Visual Content in Digital Advertising

(with R. Daviet)

Marketing Science – Revise and Resubmit (1st Round), Citation: 4 on Nov. 23

Generative AI offers vast potential for image synthesis in advertising but poses challenges in maintaining brand compatibility and performance. We address this by proposing a novel creative design process that combines generative AI with two deep Bayesian prediction models. The first model identifies potential

high-performance visuals, while the second assesses brand acceptability.

Both models undergo sequential training using optimized batches of creatives to minimize data requirements. We demonstrate the effectiveness of this approach through a field application with an outdoor activity company. The results show that our framework can generate high-performing visuals consistently, bridging the gap between the theoretical potential of GenAI and effective practical application in digital advertising.

Work in Progress:

1. Generative Priors for Conjoint Analysis: Leveraging LLMs and Variational Autoencoder for Data-Scarce Estimation

(with N. Arora)

This project addresses the limitations of Hierarchical Bayes (HB) models in data-scarce environments, such as "cold-start" problems for new technologies or B2B contexts. We propose a methodology to leverage LLMs and deep generative models to construct high-quality, informative priors.

Our approach utilizes a two-phase framework centered on a Conditional Variational Autoencoder (CVAE). In the offline phase, we employ an LLM as a "customer simulator" to generate synthetic preference data for diverse personas, training the CVAE to map textual descriptions to preference structures. In the online phase, this pre-trained model generates a bespoke prior distribution for real-world subjects. We introduce a Fully Personalized Bayesian approach, utilizing a fixed covariance derived from the simulation, which enables stable estimation even with sparse data.

Dissertation Chair

Professor Neeraj Arora
Department of Marketing, Wisconsin School of Business
Email: neeraj.arora@wisc.edu