

Xiaowei Guo

Chicago, IL | 8728186260 | viv711109@gmail.com

Education

University of Chicago, Chicago, US

Sept. 2022 – Jun. 2024

- MA in Computational Social Sciences

GPA: 3.84/4

University of International Business and Economics (UIBE), Beijing, China

Sept. 2017 – June 2021

- Bachelor of Management

GPA: 3.28/4

Research Interest

Topics: e-commerce platform, advertising, social media, user-generated content,

Methods: machine learning, statistical modeling, causal inference, experimental design

Research Experience

Research Associate, Harvard Business School

Sept. 2024 – present

Supported Projects:

The Impact of Unionization on Consumer Perceptions of Service Quality | Supervisor: Isamar Troncoso

- Data collection: Large-scale Google Maps review data and Safegraph; Starbucks store feature data via large-scale web-scraping.
- Created summary statistics and exploratory analysis.
- Analyzed Safegraph data to examine the effect of unionization on Starbucks visits and transactions using Difference-in-differences (DiD).
- Used Propensity Score Matching DiD (PSM-DiD) and Synthetic DiD to analyze the effect of unionization on Google Maps review ratings to compare with the standard DiD estimator.

Research on Price Negotiation and Influencer Agency | Supervisor: Ayelet Israeli

- Exploratively analyzed the relationship between the rolling average of video view counts and Cost per Mille (CPM) for the influencer sponsor data.
- Exploratively analyzed the relationship between the rolling average of video view counts and deal features, such as creator popularity, rounds of negotiations and negotiator.
- Performed regression analysis to explore how the deal features relate to the deal outcome and rate outcome.
- Finalized the prompt, fine-tuned GPT-4o and used the fine-tuned model to obtain rounds of price negotiation information from email thread data.

Research on Facial Images of Real Estate Agents | Supervisor: Isamar Troncoso

- Obtained face embeddings from agent profile pictures using facenet128, facenet512, VGG and dlib.
- Used Machine Learning models (logistic regression, naive bayes and random forest) to predict outcomes with and without dimension reduction using Principal Component Analysis (PCA).
- Obtained LIWC dictionary categories from the review text. Used it along with the bigram to predict agent features: beauty, age and gender. Used 10-fold cross-validation to obtain average accuracy and ROCAUC, with respect to three ML models.
- Used Latent Dirichlet Allocation (LDA) for topic modeling to cluster reviews and get major topics from the review text. Performed statistical test to find out topic-level distribution differences in beauty, age and gender when setting different numbers of topics.

The Effect of Privacy Regulations on Data Sharing Behavior | Supervisor: Eva Ascarza, Ayelet Israeli

- Collected Google Trend search index of the keyword “privacy” by state and created an interactive visualization of the time series data.
- Collected news articles with keyword search and created the weekly and monthly news article count by state data table.
- Created a summary table of user count by state out of the user receipts data of a customer engagement app, running batch R scripts on the high-performance computing cluster at Harvard Business School.

Learning from Many Experiments | Supervisor: Eva Ascarza

- Implemented network analysis with campaign user data. Created a two-mode network from the campaign-user pairs and projected it to a one-mode campaign network. Performed community detection to identify subclusters by campaign features.
- Created summary statistics and visualizations and helped with drafting the manuscript.

Research on the Impact of Virtual Tours | Supervisor: Isamar Troncoso

- Used BERT for sentiment analysis to validate the labels generated by VADER.
- Used batch GPT for image classification.
- Reviewed literature regarding the impact of photo quality, virtual tour features on platforms.

Graduate Researcher, University of Chicago

Research on Historical Trend and Current Topics in Consumer Research

Nov. 2023 – Jun. 2024

- Gathered journal metadata by searching through keyword search to extract abstracts of journal articles from the Web of Science database.
- Analyzed the trends of total publications and the publication counts with respect to keyword, and journals.
- Conducted topic modeling with both LDA and BERTopic on the abstracts of the journals. Compared the differences between the Bag-of-words approach and the embedding model. Extracted the keywords with respect to topics.
- Analyzed the trends of the publications under different topics.

Replicating Ding and Savani (2020) Paper

Mar. - Oct. 2023

- Designed and built a Qualtrics survey to conduct an online experiment replicating Ding and Savani's paper in 2020 on the effects of variability as well as people's perception of variability on within-domain behaviors as well as behaviors in unrelated domains.
- The project replicated the study 2c with public materials with an extension on the influence of the original settings on consumer behavior.
- Contributed to the write-up of the pre-registration as well as the project report reporting the method, procedures and the results and discussion of the replication.

Honors & Awards

- Excellence Award of OVAL China-Japan-Korea Business Creative Competition *July 2019*
- Comprehensive Scholarship *Dec. 2018*

Skills

- **Research Skills:** Data collection, Statistical Modeling, Causal Inference, Literature review
- **Data:** web-scraping, High Performance Computing, Database Management:
- **Machine Learning / Natural Language Processing:**
 - Decision Tree, Random Forest, SVM, naive bayes, LightGBM
 - t-SNE, PCA, K-mean, LDA, BERTopic
 - PyTorch, Tensorflow, HuggingFace, Transformers, CNN, LSTM
- **Programming:** Python programming, R programming, Bash, Qualtrics, Latex, SQL, STATA
- **Language:** Chinese, English