Sihan (Austin) Zhai

https://sihanzhai.github.io | +1 (617) 803-8262

EDUCATION

Harvard University, Cambridge, MA *May 2026*
M.A. in Business Economics, GPA: 3.83/4.0

Peking University, China *July 2023*
Master of Finance (Quantitative Finance Track), GPA: 3.89/4.0

Tianjin University, China *June 2020*
B.Eng. in Civil Engineering (Port, Channel and Coast Engineering), GPA: 3.89/4.0 (Rank: 1/61)

WORKING PAPERS

- **AI Summary Overrepresents Fake Reviews: Evidence from Amazon.**
  With Andrew T. Ching
  AI summary is designed to extract common themes from online reviews. However, fake reviews are richer in common themes. Therefore, we find that AI summary overrepresents fake reviews and may lead consumers to suboptimal decisions.
- **Platform Design with Lemons: Ranking, Certification and Endogenous Market Segmentation.**
  With Wenxiao Yang
  By developing an analytical model and a structural model, we find that through interactions with quality certifications, price-based ranking can create market segmentation that resembles second-degree price discrimination.
- **Digital Footprints and Information Misallocation: Evidence from Multi-method Studies.**
  With Michael (Xiaoquan) Zhang, Miaozhe Han, and Hongchuan Shen
  - Under Review at ***Management Science***
  - **Winner**, Best Paper Award at 2023 Asia-Pacific Marketing Academy Conference

  We find that digital footprints primarily represent users' prior informedness more than genuine interest. Therefore, algorithms designed to maximize engagement inadvertently prioritize users already familiar with a given topic, while overlooking users with higher interests.

WORK IN PROGRESS

- **Note on Identification of Stockpiling Models**
  With Andrew T. Ching and Matthew Osborne

RELATED DOCTORAL COURSEWORK

Economics PhD Field Courses *(A or A- in all of them)*

- Industrial Organization I (Static Structural, Harvard) — Ariel Pakes, Robin Lee
- Industrial Organization I (Theory, MIT) — Glenn Ellison, Jean Tirole
- Industrial Organization II (Dynamic Structural, Harvard) — Myrto Kalouptsidi
- Topics in Econometrics (Advanced Causal Inference, Harvard) — Davide Viviano

Marketing & Innovation PhD Field Courses *(A in all of them)*

- Digital Economics and Economics of AI (MIT) — Catherine Tucker
- Empirical Studies of Innovation and Digitization (Harvard) — Shane Greenstein
- Marketing Models – Fundamentals (Harvard) — Eva Ascarza
- Marketing Models – Methods (Harvard) — Eva Ascarza
- Consumer Behavior (Harvard) — Julian De Freitas

Economics Core Courses

- Principles of Econometrics (Harvard) — Phillip Heiler
- Econometric Methods (Harvard) — Dmitry Arkhangelskiy

- Microeconomic Theory I (Harvard) — Luis Armona
- Microeconomic Theory II (Harvard) — Christopher Avery, Samuel Richardson

Computer Science PhD Courses

- Quantitative Methods for NLP (MIT) — Yoon Kim, Chris Tanner, Jacob Andreas
- Deep Learning (Audited, MIT) — Phillip Isola, Sara Beery, Jeremy Bernstein

PRESENTATIONS & TALKS (* as presenter)

- **Platform Design with Lemons: Ranking, Certification and Endogenous Market Segmentation.**
  - * Conference on Information Systems and Technology (CIST), October 2025
  - * ISMS Marketing Science Conference, June 2025
- **An Empirical Study of Algorithm-Induced Online Information Misallocation.**
  - Symposium on Statistical Challenges in Electronic Commerce Research (SCECR), Jun 2024.
  - Asia-Pacific Marketing Academy Conference, Sep 2023, **Best Paper Award**.
  - Conference on Digital Experimentation (CoDE), Oct 2022.
  - Advances with Field Experiments (AFE), Jun 2022.
- **Platform Design with Manipulative Signals.**
  - * Hawaii International Conference on System Sciences (HICSS), January 2025.

INDUSTRY EXPERIENCES

ByteDance (Parent company of TikTok) *December 2019 – March 2020 & June 2020 – August 2020*

Monetization (Advertising) - Data Scientist Intern

- Through AB tests, causal inference, and descriptive analysis, analyzed large-scale consumer behavior data and logs of advertising bidding system, to detect mistakes of algorithms and decide advertising strategies
- Project 1: Proposed hypotheses that could explain the instability of the ad delivery system, tested the hypotheses with data, and finally reduced the number of extreme cases in the ad delivery system by 14%
- Project 2: Studied how advertisers and users use our products (e.g., refund rules for mistakes in the system, automation services) and gave advice to product managers

HONORS & AWARDS

Scholarship & Awards

- Techand Ecology & Environment Scholarship at Peking University, 2022
- National Scholarship at Peking University (Awarded to the top 1 student in a program), 2021
- National Scholarship at Tianjin University (Awarded to the top 1 student in a class), 2019 and 2018
- Scholarship of People's Government of Tianjin City, 2017
- Outstanding Graduates of Peking University, 2023
- Award for Academic Excellents of Peking University, 2021
- Outstanding Graduates of Tianjin University, 2020

Mathematical Contests

- National First Prize in Shenzhen Cup Mathematical Modeling Challenge (RMB 15,000), 2018
- Meritorious Winner in Mathematical/Interdisciplinary Contest in Modeling, 2018
- First Prize in Tianjin Mathematics Competition for Undergraduate Students, 2017

Research

- Best Paper Award at Asia-Pacific Marketing Academy Conference, 2023

SKILLS

- **Programming:** Python, PyTorch, C++, SQL, Mathematica, Stata
- **Languages:** English (Fluent), Mandarin (Native), Japanese (Beginner)

Appendix: Abstracts of Working Papers

AI Summary Overrepresents Fake Reviews: Evidence from Amazon

With Andrew T. Ching

Abstract: AI summary has been widely deployed to distill information from large volumes of reviews. In this research, we argue that there is an unintended consequence of AI summary – it tends to overrepresent fake reviews. Our key insights are (i) AI summary algorithms focus on extracting sentiments of common themes (i.e., keywords) from reviews, (ii) fake reviews are more similar to each other, leading to more common themes, and further a stronger influence on the overall sentiment of each common theme. To provide empirical support for our argument, we study AI summary on Amazon. We use three proxies for products which are less likely to have fake reviews: (i) Amazon-sold products, (ii) product with high review credibility grades on RateBud, (iii) products with fake reviews documented in previous research. We first confirm that fake reviews are significantly longer, more positive (even than authentic reviews with the same rating), and more similar to each other. We also find that, for products more likely to contain fake reviews, extracted keywords are mentioned by a larger number of reviews. We conjecture these mentions are dominated by fake reviews. Given that fake reviews are more positive, we further investigate if keyword's sentiment tends to be more positive for products more likely to have fake reviews – we indeed find robust evidence to support this empirical implication. Moreover, we find that compared with other products, the overall sentiments of AI summary of products with more fake reviews are relatively more positive than average ratings. Finally, we study the market-distortion effect. We find evidence that the bias of AI summary improves sales ranks of products sold by review manipulators.

Platform Design with Lemons: Ranking, Certification and Endogenous Market Segmentation.

With Wenxiao Yang

Abstract: Ranking algorithms are critical platform design mechanisms that determine seller visibility. However, existing research on platform design overlooks two key aspects: sellers' strategic signaling responses, especially the role of prices as quality signals, and how certification programs create seller heterogeneity. To address these gaps, we develop a theoretical framework examining how ranking design—specifically quality-based versus price-based—interacts with both strategic signaling and certification heterogeneity to shape market outcomes. Our analysis reveals two mechanisms. First, quality-based ranking creates a crowding-out effect: while revealing platform-held information, it inadvertently induces low-quality uncertified sellers to mimic high-quality uncertified sellers' pricing, creating pooling equilibria where prices lose informational value. Price-based ranking reverses this dynamic as low-quality uncertified sellers with cost advantages undercut competitors, generating separating equilibria where prices become informative. Second, compared to quality-based ranking, price-based ranking elevates these low-quality uncertified sellers to top positions, widening the perceived quality gap between certified and top-ranked uncertified sellers, and thus reducing competition between them. This creates endogenous market segmentation: certified sellers capture quality-sensitive consumers through premium pricing, while uncertified sellers expand market access to price-sensitive segments through competitive pricing. Empirical analysis of eBay's 2011 ranking algorithm redesign validates these predictions: after the redesign from quality-based to price-based ranking, prices of uncertified sellers declined due to intensified price competition, while prices of certified sellers slightly increased; consumers relied more on certifications and prices as quality signals; and high-quality uncertified sellers experienced significant welfare losses, while low-quality uncertified sellers and certified sellers potentially benefited.

Digital Footprints and Information Misallocation: Evidence from Multimethod Studies

With Michael (Xiaoquan) Zhang, Miaozhe Han, and Hongchuan Shen

Abstract: Online recommendation systems rely heavily on digital footprints, such as views, clicks, likes, and shares, to infer users' preferences. However, because people tend to engage more readily with content that is cognitively easier to process, these footprint metrics may reflect not genuine interests, but rather high pre-existing informedness. This paper challenges the critical assumption that digital footprints are reliable indicators of users' interests, proposing instead that they are more indicative of pre-existing informedness. Using a multi-method approach, we investigate the nature of digital footprints and their impact on current information recommendations. In a controlled laboratory environment in which users' granular behaviors are tracked via a simulated online platform, we consistently observe that footprint metrics do not align with users' interests; instead, they strongly correlate with users' pre-existing informedness. We then examine real-world online content allocation to test the predictions derived from our laboratory observations. Analysis of a large dataset comprising a battery of Facebook posts, combined with a field test on the Facebook platform, reveals a concerning pattern. Algorithms designed to exploit historical engagement footprints inadvertently prioritize users already familiar with a given topic, while overlooking users with higher interests and conversion probabilities. This misallocation leads to suboptimal information recommendations; for example, crucial health information may be disproportionately directed towards people already exhibiting high health literacy, while those who could derive greater benefit remain underserved. Our paper provides managerial implications that neglecting the origins of online footprints carries a profound impact on information accessibility and informed decision-making.