HAOLIN YANG

5307 S Hyde Park Blvd ◇ Chicago ◇ 60615

✉ haolinyang2001@uchicago.edu 📞+1 7732567911

EDUCATION

University of Chicago   **Sept. 2024 - Now**
Master of Science in Statistics

Tsinghua University   **Sept. 2020 - June 2024**
Bachelor of Arts in English **Outstanding Graduate** (GPA: 3.94/4.0)
Minor in Economics and Finance (GPA: 4.0/4.0)
Minor in Statistics (without degree) (GPA: 4.0/4.0)
Courseworks: Nonparametric Statistics, Linear Regression Analysis, Real Analysis, Econometrics, Applied Stochastic Processes

Cornell University   **July 2022 - Aug. 2022**
Summer Program, Cornell Institute for China Economic Research
Core Courses: Coding for Applied Economists (A+), Behavioral Economics (A+), Applied Econometrics (A+)

HONOR AND AWARDS

Academic Execellence Scholarship (2022-2023)
Outstanding Graduate of the School of Humanities, Tsinghua University (2024)
Outstanding Undergraduate Thesis of the School of Humanities, Tsinghua University (2024)

RESEARCH INTERESTS

Natural Language Processing, Large Language Models, Graph Neural Networks

PUBLICATIONS AND MANUSCRIPTS

1 Zhang, Y., Yang, F., Yang, H., & Han, S. (2024). Does checking-in help? Understanding L2 learners' autonomous check-in behavior in an English-language MOOC through learning analytics. *ReCALL*, 1–16. *doi:10.1017/S0958344024000144*

RESEARCH EXPERIENCES

ICL as semantic clustering plus fuzzy copy   **May. 2024 - Now**
- Discovered patterns of semantic clustering in the early layer hidden representations of different in-context demostrations inside LLMs by investigating the (non)linear separability of the representations at different layers.
- Provide empirical evidence that LLMs is approximately performing fuzzy copy among the hidden representations of icl prompts by matching the embedding matrices of different attention heads of LLMs

Undergraduate Thesis Automatic Post-Editing of English-Chinese Machine-Translated Academic Texts Jan. 2024 - June 2024
- Designed and fine-tuned an automatic post-editing model and tested its performance on ParaMed, a English-Chinese parallelcorpus that contains academic texts.
- Performed comparative statistical analyses as concerning the BLEU and TER scores of the raw machine-translated texts and the automatically post-edited texts
- Qualitatively examined the performance of the model by investigating how it helped to correct the machine translation errors and its failure scenarios

China's Import Bans and Global Waste Paper Trade   **Aug. 2023 – Nov. 2023**

- Constructed, fitted, and validated complex network models over 380000 waste paper trade flows among 151 countries between 2017 and 2022 in the Global Waste Paper Trade Network (GWPTN)
- Analyzed the impact of China's two waste paper import restriction policies in the timespan by investigating the trends in the network-level, community-level, as well as country-level indicators in the GWPTN
- Used the Naive Bayes classification algorithm to predict the links in the GWPTN in 2025 and used community detection techniques to explore the evolution of the community structure of the GWPTN till 2025

Impact of COVID-19 on Global Energy Trade Network   **April 2023 – Aug. 2023**

- Constructed the global energy trade network consisting of the top 25 energy trading countries between 2019 and 2022
- Collected information concerning the government responses to the COVID-19 pandemic in more than 200 countries between 2019 and 2022
- Quantified the impact of the pandemic on the global energy trade network by tracing the trends in the network-level and the node-level indicators
- Used panel regression techniques to examine the correlation between the COVID-related policies adopted by the countries and their power dynamics in the network

Factor Analysis and Language Education Assessment Instrument   **Feb. 2023 – June 2023**

- Wrote two chapters on factor analysis in a monograph titled *Statistical Methods in Language Education Studies*, which is currently in press
- Analyzed the similarities and differences between principal component analysis and exploratory factor analysis in theoretical and practical terms and discussed how they tend to be misused in many language education studies
- Applied the two analytical methods to the data collected by administering the Chinese Version of the Foreign Language Enjoyment Scale (C-FLES) to 260 high school English learners in China
- Used the two analytical methods to confirm a three-factor latent construct that underlies the 11 items in the scale and identified two items as being potentially redundant

Teacher-Student Interactions in Online Asynchronous Classrooms   **April 2022 – June 2022**

- Utilized Social Network Analysis techniques to model 986 instances of teacher-student interactions in chatrooms and discussion forums of an English MOOC on the XueTangX platform
- Compared the structure of the interaction networks in the first two months and the last two months of the MOOC by examining the changes in the network indicators including density, reciprocity, and degree centrality

Check-in Behaviors and Academic Performance in MOOC   **March 2021 – April 2022**

- Collected and cleaned the academic performance data of 11296 students of an English MOOC on the XueTangX platform including their check-in instances and scores in assignments and tests
- Conducted chi-square test of independence and logistic regression to reveal to what extent students' check-in behaviors affect their course completion rates
- Conducted Mann-Whitney U tests between the students who checked in regularly and those who did not check in to determine the differences in their learning behaviors

SKILLS

Statistics / Data Analysis:	Proficiency in R, Python, Stata Extensive experience in network analysis and simulation in R and Python
Languages:	TOEFL 117: Reading 30, Listening 30, Writing 30, Speaking 27; GRE 339+5.0: Verbal 169, Quantitative 170, AW 5.0; Mandarin (native), German (C1)