Yuanmo He

+44 7928549255 | y.he54@lse.ac.uk | yuanmohe.com | London

EXPERIENCE

Doctoral Researcher in Computational Social Science 09/2020 - present

The London School of Economics and Political Science (LSE)

- **Led, designed, and executed end-to-end research projects** that use advanced quantitative and computational methods (including machine learning, natural language processing, causal analysis, and social network analysis) to study socioeconomic inequality in daily behaviour and social interactions with large-scale digital trace data.
- Peer-reviewed Publication: **Yuanmo He** and Milena Tsvetkova. 2023. A Method for Estimating Individual Socioeconomic Status of Twitter Users. *Sociological Methods & Research*, https://doi.org/10.1177/00491241231168665. (**Impact Factor 6.3**; top journal in quantitative social science and sociology).
- Rewarded **Honourable mentions** at the 2023 International Conference on Computational Social Science.
- Designed and implemented **ETL** (extract, transform, load) pipelines tailored for various projects, leveraging **Python, R, SQL, Spark, Git**, Azure Cloud Computing, Twitter API, Google Geocoding API, and Facebook Marketing API.
- Employed relevant Python libraries (e.g., **Pandas, NumPy, SciPy, Matplotlib, Seaborn, NetworkX, PySpark**) to manipulate, analyse and visualise various types of data (mobile-tracking mobility data, US Census data, and data from social media platforms including Twitter, Facebook, and Yelp).
- **Proposed, implemented, and validated a new unsupervised learning method** to estimate the individual Twitter users' socioeconomic status (SES) for 3.5 million users (prior studies' benchmarks ~50,000).
- **Incorporated** insights from sociology, economics, social psychology, and consumer research to propose **a new theoretical argument** hypothesising that high SES is associated with more diverse consumption practices.
- Conducted **regression modelling**, **graph embedding**, **clustering**, and **community detection** with Python toolkits and self-developed code to test hypotheses about consumption inequality.
- **Designed natural experiments** with **matching methods** and observational data to show that Twitter users with high SES have significantly higher social capital (measured by network properties) than those with low SES.
- Implemented text analysis methods, including **sentiment analysis, topic modelling, word embedding, and large language models** to explore the differences in social interactions on Twitter between users with high and low SES.
- **Collaborated** with people from diverse academic, technical, and cultural backgrounds (e.g., physicists, biologists, ecologists, computer scientists, economists, political scientists, sociologists, and writers).
- Presented research findings at top academic conferences, including the **International Conference on Computational Social Science** (top peer-reviewed conference in computational social science) and the **Annual Meeting of the American Sociological Association** (top peer-reviewed conference in sociology).

Seminar Leader/Graduate Teaching Assistant (multiple posts), LSE 09/2021 – 04/2023

- Courses: Computer Programming, Applied Machine Learning, Introduction to Data Science and Machine Learning
- **Communicated** theoretical concepts and practical implementations of computer programming, machine learning, and natural language processing to students from **diverse academic, technical, and cultural backgrounds**.
- **Led** daily 1.5-hour computer lab teaching to two classes of 20+ students on understanding and performing **machine learning and natural language processing with R** at an intensive three-week LSE summer school course.
- Facilitated weekly/biweekly 1.5/2-hour computer lab teaching to classes of 30+ students on **computer programming and machine learning & deep learning with Python & R**.

EDUCATION AND HONORS

PhD Social Research Methods, The London School of Economics and Political Science (LSE) 09/2020 – 09/2024

- Supervisors: Milena Tsvetkova and Kenneth Benoit.
- Granted the **LSE PhD Studentship** for four years (£140,000).
- Additional training and research visit: Generative AI with Large Language Models, DeepLearning.AI, 2024; Complex Systems Summer School, Santa Fe Institute, 2023; Workshop on Inequality Across Scales, Space, Time and Domains, The Abdus Salam International Centre for Theoretical Physics, 2023.

MSc Applied Social Data Science (Distinction), LSE 09/2019 - 09/2020

- **Distinction in all modules**: Computer Programming, Data for Data Scientists, Applied Machine learning, Quantitative Text Analysis, Multivariate Analysis and Measurement, and Fundamentals of Social Science Research Design.

BSc Social Sciences (First Class Honours), University College London (UCL) 09/2016 - 06/2019

- Relevant modules: Quantitative Research Methods, Causal Analysis, Social Networks, Game Theory, Social Policy.
- Awarded the UCL Institute of Education **Faculty Medal** (the best final year undergraduate student in the faculty).
- Achieved **the highest final grade** of all graduates in the Social Research Institute.

SKILLS AND INTERESTS

Programming language & statistical software: Python, R, SQL, Spark, Git, Stata, SPSS

Python packages: NumPy, SciPy, Pandas, scikit-learn, Matplotlib, Seaborn, NetworkX, PyTorch, PySpark (non-exhaustive)

R packages: tidyverse, tm, quanteda, glmnet, tree, randomForest, e1071, igraph, statnet (non-exhaustive)

Advanced Data Analysis: machine learning, natural language processing, large language models, social network analysis, multivariate analysis, causal analysis, parallel computing, cloud computing

Languages: Chinese (native), English (bilingual).

Stand-up Comedy: Performed at diverse settings (parties, academic workshops, comedy clubs). Getting into the comedy club scene.