

E: jiani.yan@wolfson.ox.ac.uk
T: 07422587548
W: <https://vallerrr.github.io/>

Leverhulme Centre for Demographic Science,
Nuffield Department of Population Health,
University of Oxford

Department of Sociology,
University of Oxford

Summary

I am a computational social scientist dedicated to understanding health inequalities with advanced AI algorithms.

My formal training includes degrees achieved at merit and distinction level in Sociology and Demography, Computational Econometrics, and Computer Science. Currently, I am a forth-year DPhil student in Sociology at the University of Oxford. Alongside my doctoral research, I serve as a dedicated Research Assistant and consultant at the Leverhulme Centre for Demographic Science, the Department of Sociology and Nuffield College at Oxford.

My research focuses on advanced machine learning and AI, with particular emphasis on model evaluation and feature engineering

in empirical works. All of my independent work involves applying computational algorithms, supported by a strong foundation in Mathematics, Statistics, and Econometrics. I am an advanced programmer with Python, R, Stata, and JavaScript expertise. I am proficient across the entire data science workflow—from data mining and cleaning to model development, parameter tuning, interpretability, and the deployment of interactive dashboards. I also collaborate on research software projects, including contributing to the development of RobustiPy, a Python package for multi-universal analysis.

Education and Awards

DPhil in Sociology

University of Oxford

OXFORD, UK
2022-Present

- Supervised by Prof. Charles Rahal and Prof. Ridhi Kashyap
- Thesis title: “A Computational Approach to the Social Determinants of Health”
- Leverhulme Centre for Demographic Science Graduate Scholarship, 2024/25
- ESRC Grand Union Doctoral Training Partnership (DTP) Advanced Quantitative Methods (AQM) Awards
- The Julia Mead Knox Memorial Prize for Research in Demography and Population Health Winner 2025
- The Best Student Paper Winner in the 6th International Conference on Social Computing 2025
- The Best TCP Committee Member Winner in the 6th International Conference on Social Computing 2025

MPhil in Sociology and Demography

University of Oxford

OXFORD, UK
2020-2022

- Passed with Merit
- Core courses: Demographic Analysis, Life Course Research, Advanced Quantitative Methods, Research Design, Replication, Social Network Analysis.
- Dissertation is later developed and published in the American Journal of Epidemiology

MSc Financial Economics

University of Birmingham

BIRMINGHAM, UK
2018-2019

- Passed with Distinction (Overall grade:76.7)
- Specialisation in Computational Econometrics
- Nominated student representative
- Core courses: Econometrics with Financial Applications, Macroeconomics, Theory of Finance
- Dissertation title: “Combination forecasting based on American macroeconomic activity”

BSc Computer Science and Technology

Southwestern University of Finance and Economics

CHENGDU, CHINA
2014-2018

- 3rd Prize of Guanghua Entrepreneurship Competition of SWUFE (2015)
 - Best Vice President of Arts Enterprise Management Co., Ltd. (2016)
 - 2017 – 2nd Prize of the Fifteenth Science, Technology and Culture Festival of SWUFE
 - 2016-2018 - Recipient of the Innovation and Entrepreneurship Award of Computer Science SWUFE (six times).
 - Core Courses: Core Programming (C), Object-Oriented Programming (JAVA), Object-Oriented Programming (JAVAAE), Data Structures, Computer Algorithms and Analysis, Principles and Applications of Databases, Large Data Processing, Data Mining and Applications, Mobile Technology Development
-

Additional Training/ Affiliation/ Research Visits

The International Max Planck Research School for Population, Health and Data Science (IMPRS-PHDS)

Max Planck Institute for Demographic Research

ROSTOCK, GERMANY
2022-present

- Affiliated Member
- Core Courses: Demographic Methods, Digital and Computational Demography, Population Health
- Research Visit hosted by Dr.Marcus Ebeling, Leader of Max Planck Research Group on Medical Demography

ESRC Center For Care UKRI, NIHR and university collaboration • Affiliated Member	SHEFFIELD, UK 2022-present
Office of Population Research Princeton University • Research Visit hosted by Matthew Saganik, Alexander Stewart 1886 Professor of Sociology.	PRINCETON, US 2025
Helsinki Institute for Demography and Population Health University of Helsinki • Research Visit hosted by Pekka Martikainen, Professor of Demography.	HELSINKI, FINLAND 2024
Summer Institute in Computational Social Science University of Oxford • Core Courses: Natural Language Processing, Machine Learning, Computer Vision, Web-scraping, Geospatial Analysis, Data Visualisation.	OXFORD, UK 2021

Research Experience

Research Assistant at the Department of Sociology, University of Oxford

Digital Gender Gaps

2021-present

Project Leader: Ridhi Kashyap, Professor of Computational Social Science, University of Oxford

- Harmonising Chinese prefecture-level development indicators (such as education and medical resources)
- Facebook Marketing API Collector: Built a program to retrieve daily and monthly active audiences with different age groups and device information from Facebook Marketing API, worldwide.
- Dockerised PostgreSQL database and Flask APP construction.
- Project Pipeline construction, from downloading, preprocessing and analysing the daily collected data to uploading the results to a database linked to the back-hand of a web application.
- Database daily maintenance.
- Sub-national data collection and customised tessellation for Low- and Middle-Income Countries.
- Dashboard development for output quality check.

Research Assistant at the Leverhulme Center for Demographic Science, University of Oxford

A Scientometric Review of 20 years of UK Biobank

2025-present

Project Leaders:

Melinda Mills, Nuffield Professor of Sociology, Director of LCDS, University of Oxford;

Charles Rahal, Associate Professor of Social Data Science, University of Oxford

- Maintain and update the dimension research paper and patent scraper.
- Design and develop research on the UK Biobank author collaborations overtime.
- Design and develop research on the patents linked to the published UK Biobank and explore the impact.

GWAS Diversity Monitor

2021-2022

Project Leaders:

Melinda Mills, Nuffield Professor of Sociology, Director of LCDS, University of Oxford;

Charles Rahal, Associate Professor of Social Data Science, University of Oxford

- Dockerised database construction.
- Maintenance of HTML and content.
- Debugging and enabling the function of Google Analytics on the website to track each component's visits.
- Debugging download function.
- Debugging and rewriting various figure functions to ensure it work with all possible drawing pathways.

A Scientometric Review of Population Studies

2021

Project Leaders:

Melinda Mills, Nuffield Professor of Sociology, Director of LCDS, University of Oxford;

Charles Rahal, Associate Professor of Social Data Science, University of Oxford

- Manually reviewed over 800 pieces of population-related literature and classified them into topics.

COVID-19 Vaccine Deployment Project

2020

Project Leaders:

Melinda Mills, Nuffield Professor of Sociology, Director of LCDS, University of Oxford;

Charles Rahal, Associate Professor of Social Data Science, University of Oxford

- Harmonising and filtering the literature on human vaccination in a corpus of over 20,000 documents.
- Contributing to the creation of a map of vaccination attitudes (effective, important and safe)

Research Assistant at Nuffield College, University of Oxford

Gun Violence Research

2021-2025

Project Leader: David Kirk, Professor of Criminology, University of Pennsylvania

- Built a scraper to recursively and repeatedly fetch information on guns for sale from armslist.com.
- Deployed the scraper to the Amazon Web Services and set up a dockerised PostgreSQL database.
- Maintenance of the scraper: building a monitoring system to supervise daily data collections.

Teaching Experience

Leverhulme Center for Demographic Science, University of Oxford

OXFORD, UK

Co-convenor: Python for Health, Economic, and Social Science

2026

- Facilitated afternoon laboratory workshops throughout the course for students, focusing on introducing Python as an essential tool for statistical analysis from scratch.

Central Bank of Colombia

CARTAGENA, COLOMBIA

Co-lecturer: Python for Population Projection with PyCCM

2025

- Facilitated laboratory workshops throughout the course for senior researchers in coding Python for demographic researches from scratch.

Department of Politics and International Relations, University of Oxford

OXFORD, UK

Co-lecturer: Intro to Statistics

2023-2024

- Facilitated two R laboratory workshops throughout the course for postgraduate students, focusing on introducing R as an essential tool for statistical analysis.

Publications and Working Papers

Published Works

- Valdenegro, D., **Yan, J.**, Dai, D., Rahal, C. (2026), Introducing RobustiPy: An efficient next generation multiversal library with model selection, averaging, resampling, conditionally accepted at *Patterns*. <https://arxiv.org/abs/2506.19958>
- **Yan, J.**, Rahal, C. (2025) On the unknowable limits to prediction. *Nature Computational Science* 5, 188–190. <https://doi.org/10.1038/s43588-025-00776-y>
- **Yan, J.** (2025). Revisiting the social determinants of health with explainable AI: a cross-country perspective. *American Journal of Epidemiology*, 195(3), 681-688. <https://doi.org/10.1093/aje/kwaf205>
- Breen, C. F., Fatehkia, M., **Yan, J.**, Zhao, X., Leasure, D. R., Weber, I., Kashyap, R. (2025) Mapping subnational gender gaps in internet and mobile adoption using social media data. *Proceedings of the National Academy of Sciences*, 122(42), e2416624122. <https://doi.org/10.1073/pnas.2416624122>
- Leasure, D.R., Kashyap, R., Rampazzo, F., Dooley, C.A., Elbers, B., Bondarenko, M., Verhagen, M., Frey, A., **Yan, J.**, Akimova, E.T., Fatehkia, M., Trigwell, R., Tatem, A.J., Weber, I. and Mills, M.C. (2023). Nowcasting Daily Population Displacement in Ukraine through Social Media Advertising Data. *Population and Development Review*. <https://doi.org/10.1111/padr.12558>

Policy Reports

- Mills, M., Rahal, C., Brazel, D., **Yan, J.**, and Gieysztor, S. (2020). COVID-19 Vaccine Deployment: Behaviour, ethics, misinformation and policy strategies. *London: The Royal Society & The British Academy*.

Software and Tools Development

- Leasure DR, **Yan, J.**, Bondarenko M, Kerr D, Fatehkia M, Weber I, Kashyap R. (2023). Digital Gender Gaps Web Application, v1.0.0. Zenodo, GitHub. doi:10.5281/zenodo.7897491. <https://github.com/OxfordDemSci/dgg-www>.
- Valdenegro, D., **Yan, J.**, Dai, D., Rahal, C. (2025). *RobustiPy: A Python Package for Robustness Checks in Social Science Research*, v0.1.0. Zenodo, GitHub. doi:10.5281/zenodo.15700698. <https://github.com/RobustiPy/robustipy>

Working Papers

- Rahal, C., **Yan, J.**, M., Verhagen, (2026). From a Single Seed Grows a Forest of Uncertainty, in progress/preparation for submission to leading journals and conferences.
- **Yan, J.** (2026). Discovering Prevalent Chronic Disease Profiles with Advanced Clustering Methods in UK Biobank, in progress/preparation for submission to leading journals and conferences.
- **Yan, J.** (2026). An Item Response Theory Approach to Model Multimorbidity Complexity Trajectory and the Role of Social Determinants in UK Biobank, in progress/preparation for submission to leading journals and conferences.

- **Yan, J.**, Li, J., Breen, C.F., Kashyap R. (2026). Nowcasting global digital gender gaps using social media data, in progress/preparation for submission to leading journals and conferences.
- Wang, P., Rahal, C., **Yan, J.**, Kanopka K., Domingue, B. (2026). Mastering Model Fit: Applications of the InterModel Vigorish for Interpretable Machine Learning, in progress/preparation for submission to leading journals and conferences.

Conferences

- International Conference on Social Computation 2025 SHANGHAI, CHINA
Oral presentation: Discovering Prevalent Chronic Disease Profiles with Advanced Clustering Methods in UK Biobank
- Demystifying Machine Learning workshop 2024 by MPIDR ROSTOCK, GERMANY
Oral presentation in session Machine Learning Approaches for Population Research, presenting work: Death Predictions from Social Determinants: Holistically and Precisely with Explainable AI
- PAA 2025 Annual Meeting WASHINGTON D.C., US
Poster Session, presenting works: Discovering Prevalent Chronic Disease Profiles with Advanced Clustering Methods; Nowcasting Global Digital Gender Gaps Using Social Media Data
- European Population Conference 2024 EDINBURGH SCOTLAND, UK
Oral presentation in session Machine Learning Approaches for Population Research, presenting work: Death Predictions from Social Determinants: Holistically and Precisely with Explainable AI
- PAA 2024 Annual Meeting COLUMBUS, OHIO
Poster Session, presenting work: Predicting death Holistically and Precisely with Explainable AI
- 9th International Conference on Computational Social Science 2023 COPENHAGEN, DENMARK
Poster Session, presenting work: Predicting death Holistically and Precisely with Explainable AI

Review Experience

- Scientific Report (X1)
- American Journal of Epidemiology (X1)
- International Conference on Social Computing (16 Submissions)

Extra-curricular Activities

- Metrics and Models UK
Core co-organising member 2025-present
• Co-organising bi-weekly online seminar on the metrics and models used across health and social data science
- International Conference on Social Computing UK
Core co-organising committee member 2025-present
• Co-organising the international conference in both Shanghai (2025) and Oxford (2026)
• Serving as TCP member for Shanghai 2025 conference, review submitted papers (Best TCP committee member winner)
- European Association of Population Study Netherlands
Steering Committee Member 2024-2025
• Organising events for and with PhD students and early career researchers in Europe and in collaboration with EAPS.
• Organising early career/student events and webinars.

Referees

Referee 1

Prof. Charles Rahal
Associate Professor in Data Science and Informatics
Leverhulme Centre for Demographic Science,
Nuffield Department of Population Health,
University of Oxford.
Email: charles.rahaldemography.ox.ac.uk

Referee 2

Prof. Ridhi Kashyap
Professor of Demography and Computational Social Science
Department of Sociology,
University of Oxford.
Email: ridhi.kashyap@sociology.ox.ac.uk

Referee 3

Prof. David Kirk
Professor and Department Chair of Criminology
Department of Criminology,
University of Pennsylvania.
Email: davekirk@sas.upenn.edu

Referee 4

Prof. Casey Flather Breen
Assistant Professor
Department of Sociology,
University of Texas
Email: casey.breen@austin.utexas.edu