

Summary

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

My formal training includes degrees achieved or expected at distinction level in Sociology and Demography, Computational Econometrics, and Computer Science. I am presently a first-year DPhil student in Sociology at the University of Oxford, and a persistent Research Assistant and consultant within the Leverhulme Centre for Demographic Science.

I specialise in advanced machine learning, model evalua-

tion, and feature engineering. All of my individual research involves machine learning algorithms. More broadly, this is made possible through my solid foundations in Mathematics, Statistics and Econometrics. I am an advanced programmer in multiple languages, including Python, R, Stata and JavaScript. I am proficient across all parts of a data-scientist research pipeline, including data mining, data cleaning, model building, parameter tuning, model interpretation, and, finally, the construction of interactive dashboards.

Education

DPhil in Sociology

University of Oxford

OXFORD, UK

2022-Present

- Fully Funded by ESRC Grand Union Doctoral Training Partnership (DTP) Advanced Quantitative Methods (AQM) Awards

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: "How Predictable is Death?"

MSc Financial Economics

University of Birmingham

BIRMINGHAM, UK

2018-2019

- Passed with Distinction
- 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 Economics

CHENGDU, CHINA

2014-2018

- 2015 -- 3rd Prize of Guanghua Entrepreneurship Competition of SWUFE
 - 2016 -- Best Vice President of Arts Enterprise Management Co., Ltd.
 - 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 (JAVAEE), Data Structures, Computer Algorithms and Analysis, Principles and Applications of Databases, Large Data Processing, Data Mining and Applications, Mobile Technology Development
-

Additional Training / Affiliation

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

Max Planck Institute for Demographic Research

ROSTOCK, GERMANY

- Affiliated Member
- Core Courses: Demographic Methods, Digital and Computational Demography, Population Health

ESRC Center For Care

UKRI, NIHR and university collaboration

UK

2022-present

- Affiliated Member

- Core Courses: Natural Language Processing, Machine Learning, Computer Vision, Web-scraping, Geospatial Analysis, Data Visualisation.
-

Postgraduate Dissertation

Research Experience

University of Oxford

OXFORD, UK
2020-2022

Research Assistant at the Leverhulme Center for Demographic Science

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 and click-through rates.
- 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 choropleth map of vaccination attitudes (effective, important and safe)

Research Assistant at the Nuffield College

Gun Violence Research

2021-now

Project Leader: David Kirk, Professor of Sociology, University of Oxford

- 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.

Research Assistant at the Department of Sociology

Digital Gender Gaps

2021-now

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.

Crime Research

2022

Project Leader: Federico Varese, Head of Department, Professor of Criminology, University of Oxford

- Built a scraper to fetch the publicly available profile of criminals.
 - Harmonise the data to the format that facilitates qualitative research.
-

Teaching Experience

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

- Mills, M., Rahal, C., Brazel, D., **Yan, J.**, & Gieysztor, S. (2020). COVID-19 Vaccine Deployment: Behaviour, ethics, misinformation and policy strategies. *London: The Royal Society & The British Academy*.
- 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>.
- 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>
- **Yan, J.**, Rahal, C. On the unknowable limits to prediction. *Nature Computational Science* 5, 188–190 (2025). <https://doi.org/10.1038/s43588-025-00776-y>

Working Papers

- Rahal, C., **Yan, J.** (2022). ‘From a Single Seed Grows a Forest of Uncertainty.’, in progress/preparation for submission to leading journals and conferences.
- Rahal, C., Valdenegro, D., **Yan, J.** (2024), ‘Introducing Robustify: An efficient multiversal library with model selection, averaging, and out-of-sample analysis’, in preparation for submission to leading journals and conferences.
- Breen, C. F., Fatehkia, M. **Yan, J.**, Zhao, X. Leasure, D. R., Weber, I., Kashyap, R. (2024) ‘Mapping subnational gender gaps in internet and mobile adoption using social media data’
- **Yan, J.** (2023), ‘Revisiting the social determinants of health with explainable AI: a cross-country perspective’, under review.
- **Yan, J.** (2024), ‘Bridging Social Factors to Pathological Outcomes: A Deep-Learning Pseudo Imaging based Approach’, in progress/preparation for submission to leading journals and conferences.

Conferences

- 9th International Conference on Computational Social Science (IC2S2) COPENHAGEN, DENMARK
Poster Session, presenting work: Predicting death Holistically and Precisely with Explainable AI
- PAA 2024 Annual Meeting COLUMBUS, OHIO
Poster Session, presenting work: Predicting death Holistically and Precisely with Explainable AI
- 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

Languages and Skills:

- | | | | | |
|----------|--------------|---------|----------|-----------------------------------|
| • Python | • Java | • IELTS | • Eviews | • Japanese (N3) |
| • R | • JavaScript | • Stata | • HTML | • L ^A T _E X |

Extra-curricular Activities

European Association of Population Study

Netherlands

Steering Committee Member

2024-2026

- 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.

Professional History

Campus President, Chengdu Arts Enterprise Management Co., Ltd.

Chengdu, China

Campus President 2016

- Supervised operations.
- Took in charge of the management of SWUFE intra-campus company operation.
- Acted as the main coordinator of musical instrument training and operation.
- Attracted and encouraged engagement and participation.

