XIAOYI (MIMI) CHEN

mimixiaoyichen@gmail.com • + 1 407-233-8925 • https://mimixiaoyichen.com/

EDUCATION

Oxford University, Oxford Internet Institute *Oxford, UK*
MSc in Social Data Science *June 2025*
New York University, College of Arts and Sciences *New York, NY*
B.A. in Data Science, Minor in Linguistics | GPA: 3.8/4.0 *(cum laude)* *May 2024*

SKILLS

Programming Languages:	Python, Java, R, JavaScript, HTML/CSS
Big Data & Machine Learning:	SQL, Python (nltk, scikit-learn, pytorch, hugging face), LUIS
Other:	Microsoft Suite, & Azure, Figma, Adobe Suite, Git, LaTeX

PROFESSIONAL EXPERIENCES

PricewaterhouseCoopers (PwC) *New York, NY*
Technology Consulting Intern - Cloud & Digital *June 2023 - August 2023*
- Built and deployed three PowerBI dashboards to manage data for a cross-functional team of 350+ project managers and data scientists, resulting in an 80% reduction in required meeting times
- Developed and delivered presentations for client projects in generative AI applications, providing use cases and insight into the power of machine learning in leading industries such as food and media

OTIS Elevator Worldwide *New York, NY*
Data Analytics Intern *June 2022 - August 2022*
- Developed a self-guided chatbot using Microsoft Azure and LUIS to be utilized internally by 10,000+ users across 1400 branches, reduced the time spent troubleshooting common issues by 50% worldwide
- Collaborated with the lead of Otis' Project Management team to develop a user persona for viewing financial metrics and objectives, focusing on data granularity and data-driven insights for PMs

NYU Center for Social Media and Politics *New York, NY*
Research Assistant *October 2021 - April 2022*
- Analyzed social media content across 10 platforms and conducted research on current events to facilitate research into politicians' online behaviors and the subsequent online and offline effects
- Collected and managed datasets exceeding 30,000 instances, including data on politicians' social media activities to assess the influence of social media on political communication during campaigning

PROJECTS

Quantifying Ideological Homophily in the Popular Political Podcast Ecosystem *Spring 2025*
- Modelled guest-host interactions of the top 100 political podcasts through a bipartite network analysis, leveraging community detection, attribute assortativity, and path-based metrics to measure ideological clustering in the podcast ecosystem and examine patterns in online elite political discourse
- Demonstrated significant ideological homophily through permutation tests of network metrics in the projected unipartite graph and utilised bipartite modularity maximisation to identify community formation

Sexism Detection in the 2024 U.S. Presidential Election with DeBERTa *Spring 2025*
- Finetuned deberta-xlarge on labelled data from Reddit and Gab to detect binary and granular sexism
- Pseudo-labeled Twitter posts using an entropy-aware MC Dropout strategy to identify sexist content during the 2024 U.S. Presidential Election, enabling large-scale quantification of online sexism

Heterogeneity in Drivers of Higher Education Attainment: Evidence from 21st-century China *Fall 2024*
- Conducted comprehensive analysis of panel data from 2005 to 2019 to evaluate the impact of various metrics on new higher education enrollments across Chinese provinces using a fixed-effects model
- Identified urbanisation and regional GDP as the greatest influence on education outcomes in 21st-century China as well as heterogeneous effects across economic regions and temporal stages

Anime Genre Classification with Convolutional Neural Networks *Winter 2023*
- Built a custom Convolutional Neural Network to predict Anime genres from posters and compared model performance to a transfer learning model from ResNet50, achieving a Hamming Loss of 0.087
- Utilized automated machine learning (AutoML) to evaluate its effectiveness in hyperparameter tuning with AutoKeras to experiment with 10 different model architectures and hyperparameter sets

The Food Delivery Dilemma: Classification & Evaluation of Real vs. Ghost Kitchens *Spring 2023*
- Developed a classification tool to analyze linguistic patterns in menu item descriptions from online food delivery platforms using OpenAI's text-embedding-ada-002 and cosine similarity
- Demonstrated the potential of using NLP-based methods for tackling the growing concern about transparency and quality in the online delivery food industry

Fundamentals of Machine Learning Capstone Project: Music Genre Classification *Spring 2023*
- Leveraged features from 50,000 songs from the Spotify API to predict the genre using classification
- Applied dimensionality reduction and k-means clustering to determine appropriate amount of features to retain and optimal number of clusters for XGBoost classification, achieved ROC-AUC score of 0.902

EXTRACURRICULAR ACTIVITIES

MSc Student Representative *Oxford, UK*
Equality, Diversity, Inclusion Committee, Oxford Internet Institute *October 2024 - Present*
- Student representative to department committee providing an open forum for discussion of equality and diversity issues, advising department leadership on student issues and making event recommendations
- Attend committee meetings to deliberate on policies and initiatives aimed at fostering an inclusive environment for department members, bridging communication between student body and leadership