

# XIN YE

xy2509@columbia.edu ♦ +1 (626)766-9947

## EDUCATION

---

### Columbia University

M.S. in Data Science

Expected Dec 2022

*New York, NY*

### University of California, Davis

B.A.S in Statistical Data Science & Economics Analysis

Overall GPA: 3.85/4.00

June 2021

*Davis, CA*

## PROFESSIONAL EXPERIENCE

---

### Team Leader for Capstone Project

*Columbia University*

Jan 2022 - present

*New York, NY*

**Supervisor:** Dr. Sanghamitra Dutta, AI researcher at JPMorgan Chase & Co.

**Topic:** Quantifying Contributions of Different Features to Unfairness in AI

- Implement SHAP to quantify different types of fairness(e.g. demographic parity, equalized odds, counterfactual fairness, etc.) in different machine learning models, such as Logistic Regression, Random Forest, XGBoost, LightGBM, and CNN
- Conduct empirical experiments involving hyper-parameter tunings and model selections to discover patterns between model performance and feature's contribution to unfairness in different ML models
- Brainstorm with teammates, present results to mentors and distribute tasks weekly

### Student Researcher

*Strohmer Lab, UC Davis*

Jan 2020 - Sept 2020

*Davis, CA*

**Supervisor:** Professor Thomas Strohmer

**Topic:** Privacy-Preserving Synthetic Transportation & Numeric Images Generation

- Constructed GANs with autoencoder to synthesize mathematical certifiable similar distributed transportation and numeric images without revealing privacy on Google Colab using Pytorch
- Built entropy grader as well as using Frechet Inception Distance (FID) to evaluate the performance of GANs
- Improve the performance of GANs for different testing data sets such as MNIST with an accuracy of 98.7%, and CIFAR-10 using Deep Convolutional GAN(DCGAN), Wasserstein GAN(WGAN), Conditional GAN(CGAN), and Auxiliary Classifier GAN(ACGAN) by hyperparameter tuning

### Data Assistant

*School of Education, UC Davis*

June 2019 - June 2020

*Davis, CA*

**Supervisor:** Robin Martin

- Developed database system for more than 30 budget proposals, 200 research artifacts, and 200 survey responses and offer administrative support for projects and educational conferences
- Compiled economic and demographic data of more than 30 target schools into Google Spreadsheet

## SKILLS & LANGUAGES

---

### Technique Skills

Python, MySQL, R, scikit-learn, Tensorflow, PyTorch, D3.js, R shiny, L<sup>A</sup>T<sub>E</sub>X, Tableau, BigQuery, HDFS, GCP, AWS, Spark

### Data Analysis Skills & Specialities

Exploratory Data Analysis, Machine Learning Models & Techniques (Supervised Learning: Linear Regression, Random Forest, Decision Trees, SVM; Unsupervised Learning: clustering, PCA; Gradient Boosting, CNN, GAN, NLP, etc.), Database System, Statistics