# Yuhan (Alison) Yao

yy2564@nyu.edu | +86 18939704386 | Saadiyat Island, Abu Dhabi, United Arab Emirates
Personal Website | LinkedIn | GitHub | Medium | Kaggle

## EDUCATION

**New York University Shanghai,** Shanghai, China — Expected graduation date: 05/2022
Bachelor of Science in *Data Science with a concentration in AI*; Minor: *Math* — Cumulative GPA: 3.93/4.0

Study Abroad: **New York University Abu Dhabi,** United Arabic Emirates — 09/2021 - 12/2021
Study Abroad: **New York University New York,** New York City, USA — 01/2022 - 05/2022

## RESEARCH & PUBLICATION

**Optimization of NYU Shanghai Shuttle Bus Schedule**, Shanghai, China — 05/2021 - 09/2021

*Undergraduate Student Researcher (Received Dean's Undergraduate Research Fund); Advisor: Prof. Zhibin Chen* — GitHub
- Applied an enhanced Genetic Algorithm using Python to solve a black-box optimization problem and devised a new and improved shuttle bus schedule, which reduced cost by up to 20% while satisfying students' demand
- Formulated a real-life vehicle scheduling problem into variations of Spatio-temporal networks, constructed a non-closed form objective and various constraints, and presented research outcome as a user-friendly Flask web app
- Led a team of 2 students and 1 shuttle service supervisor to collaborate and communicate with advising professor

**History Beyond: Approaches to Messy Digitized Archival Documents**, Shanghai, China — 08/2020 - 09/2020

*Research Assistant; Advisor: Prof. Heather Ruth Lee* — Project Website | GitHub
- Implemented Optical Character Recognition using Python OpenCV and Google Tesseract to recognize English words in ancient fonts, digitalizing and preserving historical documents
- Utilized Python Pandas package to wrangle and organize tabular data with 18,000+ entries in Chinese Restaurant Database
- Presented research outcome to NYU Shanghai Chancellor, Provost, Dean and Professors

**Open Source Swarm Intelligence Robotics Research**, Shanghai, China — 03/2019 - 12/2019

*Research Assistant; Advisor: Prof. Rodolfo Cossovich* — Hackaday
- Designed, 3D-printed, tested and compared prototypes for infrared perception of robots "Swarmesh", building scalable and decentralized swarm intelligent robots from scratch
- Published paper Framework for Present Swarm Robotic Systems and New Implementations to Increase Scalability
- Presented paper at SWARM 2019: The 3rd International Symposium on Swarm Behavior and Bio-Inspired Robotics in Japan

## WORK EXPERIENCE

**Coopsight, LLC**, Shanghai, China — 04/2020 - Present

*Data Scientist* — 08/2021 - Present
- Tested business hypotheses and product-market fit by cold-calling 50+ people and interviewing 20+ potential customers and industry experts to propose data-driven solution for product iteration
- Designed and tested data solution feasibility through extensive research and web scraping data collection methods

*Software Developer* — 04/2020 to 08/2021
- Designed, implemented and tested multiple web pages using React.js and Tailwind.css, including embedded Google map, company searching, user matching, enhancing user experience
- Managed and maintained data extraction and modification of Firebase NoSQL database, ensuring stable and smooth data connection between frontend and backend

**PayPro Global**, Remote at New York Office — 02/2021 - 05/2021

*Data Science Intern*
- Utilized Python to implement K-means and Hierarchical Clustering to transform customer recency, frequency, and monetary values, which segmented customers into 3 target groups and prepared 8 features for prediction model
- Constructed, tested and finetuned customer lifetime value prediction model with 85%+ accuracy using XGBoost, which assisted marketing team to refine their customer target strategy
- Created and designed Power BI data visualization report featuring 16 plots on webpage template performance, enabling frontend team to debug hidden template errors and optimize template functionality
- Presented results to CEO and team leader and wrote requested executive summary detailing model mechanism and marketing strategy for senior leadership

**Shanghai Hyron Software Co., LTD**, Shanghai, China 06/2020 - 08/2020

*Artificial Intelligence Intern*

▪ Extracted structural information from document photos using OpenCV, trained a 95%+ accurate CRNN model to recognize typed numbers and 7000+ Japanese characters, and implemented facial verification on ID photos using Insightface
▪ Utilized Python to implement an automated pose detector of abnormal behavior by using YOLOv5 and Resnet18
▪ Designed application scenarios and tested VGG and Resnet on classification performance and YOLOv5 on objective detection performance

## SCHOLARSHIPS & HONORS

| | |
|---|---|
| Good Wood Global Scholar, NYU Shanghai | 2021 - 2022 |
| Dean's Undergraduate Research Fund, NYU Shanghai | 05/2021 |
| Dean's List 2019-2021, NYU Shanghai | 2019 - 2021 |
| Recognition Award 2019-2020, NYU Shanghai | 2019 - 2020 |
| Presentation at Academic Conference (PAC) Grant, NYU Shanghai | 11/2019 |
| Global Quintessence Scholarship ($39,000 awarded for top 1%), NYU Shanghai | 2018 - 2022 |

## ACTIVITIES

| | |
|---|---|
| **Technical Writer**, Towards AI (Medium Publication) | 07/2021 - Present |
| **Open-source Dataset Contributor**, Kaggle | 08/2020 - Present |
| **Senior Student Worker**, Career Development Center, NYU Shanghai, Shanghai, China | 09/2019 - 05/2020 |
| **Treasurer**, MUBOtics Club, NYU Shanghai, Shanghai, China | 03/2019 - 12/2019 |
| **Marketing Leader of Yanghuo Project**, Match Edu (NGO), Shanghai, China | 06/2019 - 08/2019 |

## DATA SCIENCE COMPETITION & PROJECTS

**1st Place in Eleme Delivery Analytics Under COVID-19**, Kaggle Data Science Competition Kaggle GitHub
Built, compared, and optimized various Machine Learning models such as logistic regression and XGBoost to pinpoint relevant factors conducive to predicting couriers' decisions and behavior. Narrowed down the error of delivery time prediction to around 1.5 minutes.

**Bechdel Test: Comparing Female Representation Metrics in Movies**, Human-centered Data Science Final Project GitHub
Analyzed Bechdel Test scores of 9,300+ movies over about 150 years and summarized the trend of female representation evolution using data visualization; Compared the Bechdel Test with other metrics such as director gender, cast and crew female ratio through statistical analysis; Use Thick Data to qualitatively criticize the limitations of the Bechdel Test and provide improvement advice on creating a superior metric.

**NL2SQL: BERT-based Model for SQL Generation**, Natural Language Processing Final Project GitHub
Designed and built a BERT-based slot-filling classification model that converted questions in Chinese into SQL statements, which enabled non-programmers to interact with SQL databases effortlessly in Q&A scenarios.

**SAC Bias Reduction: Clipped Double Q vs Multi-Step Method**, Reinforcement Learning Final Project GitHub
Tested and compared clipped double Q, multi-step method, and the combination of them in terms of SAC bias reduction efficacy for algorithm optimization.

**Chinese Traffic Sign Recognition Based on Annotated Street View Images**, Machine Learning Final Project GitHub
Created and trained a self-designed artificial neural network on 6000+ images and tested it against VGG16 and Resnet50 to accurately classify 58 categories of Chinese traffic signs.

**Video Streaming Platform: Which Has Better Shows**, Regression & Multivariant Data Analysis Report GitHub
Utilized Minitab to conduct a two-way ANOVA multiple regression analysis on ratings of different genres from major video streaming platforms such as Netflix, Amazon Prime, Hulu, Disney+, and Apple TV+, which identified the relationship between ratings and two categorical variables: platform and genre.

**Online Air Ticket Reservation System**, Databases Final Project GitHub
Designed and implemented a relational database using MySQL and created a frontend user interface using HTML, CSS, Bootstrap and JavaScript to connect with the backend Flask app.

## SKILLS

**Data Science:** Python (NumPy, Pandas, Matplotlib, Plotnine, Scikit-learn), R (ggplot2, tidyverse), SQL, Firebase NoSQL
**AI:** Python (Pytorch, OpenCV), Machine Learning, Computer Vision, Reinforcement Learning, Natural Language Processing
**Software Development:** HTML, CSS (SASS/SCSS, Tailwind), JavaScript (React.js), Python Flask
**Language:** Chinese (Native), English (Professional Work Proficiency, TOEFL: 112/120, GRE: 328/340)