Yuan Jack Yao

404-528-6519 | yyao411@gatech.edu | LinkedIn | GitHub | Portfolio

Education

Georgia Institute of Technology **Aug 2025 – Dec 2026**

MSc Analytics @ ***School of Computational Science and Engineering*** *Atlanta, Georgia*

- GPA 4.0/4.0: **Conversational AI, Natural Language, Machine Learning, ML in Computational Biology**, Data & Visual Analytics, Computing for Data Analysis, Nonparametric Data Analysis, Business Analytics Practicum

National Taiwan University **Sep 2014 – Jun 2018**

Bachelor's Degree in Economics *Taipei, Taiwan*

Work Experiences

Walmart **Jun 2026 – Aug 2026**

Data Scientist III Intern @ Sam's Club AI Acceleration Team *Bentonville, Arkansas*

- Deployed an **automated hourly snapshot pipeline** with **Docker** and **Kubernetes**, integrating **asynchronous processing** that reduced execution time by **50%**, and fully covering **1,000+** cameras' snapshots into **Azure**.

SUEZ **May 2022 – Sep 2024**

Data Scientist @ Smart & Environmental Solutions (SES) China *Shanghai, China*

- Streamlined an **NLP-based Keras** classifier for **imbalanced** SCADA alarms at **82% production accuracy**.
- Speeded up classifier application to **1 second runtime per HTTP request** in **Linux Services**.
- Developed a **DBSCAN** anomaly detector to send **predictive maintenance alarms** for air compressors.
- Anomaly detector reached **70% production F1 score** with timeliness of **2 weeks earlier than actual events**.
- Applied **Joblib** on Linux and **saved 20% processing time** of 12 air compressors' clustering.
- Designed an **automated MySQL ETL pipeline** with conditional deletion on minutely time series.

Open Source Contributions

LeetCode Wiki (36.2k Stars, 9.4k Forks)

- Authored **8 optimal time and space solutions with tutorials** for **hard** problems. Selected merged PRs include:
- **No.2454 Next Greater Element IV**: Achieved $O(n)$ time complexity by **double stacks** in 2024, and merged PR in 2026, reducing from original solution's global sorting and ordered set of $O(n \log n)$ time.
- **No.3430 Maximum and Minimum Sums of at Most Size K Subarrays**: Delivered the repository's first tutorial for No.3430, achieving both time and space complexities at $O(n)$ via **deques**.
- **No.1793 Maximum Score of a Good Subarray**: Optimized $O(1)$ space complexity with **greedy two pointers**, reducing from original solution's stack of $O(n)$ space.

Personal Projects

JOB: Jack's Online Blackjack *Web Game, Python, FastAPI, JavaScript, CSS*

- Developed a full-stack online Blackjack game with casino-style UI via FastAPI backend & vanilla JavaScript frontend.
- Implemented Macau style complex rules: up to 3 splits, double after split, early pay option, and insurance.
- Constructed CI/CD pipeline via GitHub Actions and Railway, enabling automatic deployment on every git push.

SkyHorse *Bilingual Technical Blog, Data Structures, Algorithms*

- Authored in-depth tutorials for 20+ LeetCode hard problems, delivering optimized solutions with complexity analysis.
- Each tutorial provides algorithmic intuition, step-by-step breakdown, with implementations in both Python and C++.
- Navigatable catalog spans stacks/queues, sliding window, prefix sum, hash table, heaps, greedy/DP, DFS/BFS.
- Entire website language supports English (default) and Chinese, which are switchable per user preference.

EcoTeller *PyTorch, HuggingFace, Supervised Fine-tuning, LoRA*

- Fine-tuned Qwen2.5-3B-Instruct via HuggingFace to extract and quantify environmental issues from news titles.
- Designed sparse encoding pipeline to convert text vectors into structured prompts for Qwen's chat template.
- Built batch inference system with custom parser to decode model-generated sparse lists into issue-severity vectors.

Technical Skills

Gen AI & ML: LLMs Fine-tuning, Distillation, Prompt Engineering, PyTorch, Hugging Face, Keras, TensorFlow
Automation & CI/CD: GitHub, Bitbucket, Linux, Docker, Kubernetes
Cloud & Data Platforms: Microsoft Azure, Amazon Web Service, Google Cloud Platform
Programming & Databases: Python, MySQL, PostgreSQL, ETL Pipelines, C++
Frontend & Visualization: FastAPI, JavaScript, CSS, HTML, Flask, Web Scraping