

Yiming (Polo) Huang

Taipei, Taiwan | +886-919-120-987 | toy042897@gmail.com | www.linkedin.com/in/jsbao

6+ years of experience in **Data Engineering**, Backend Development, and **MLOps**. Specializing in building internal backend and data platforms. Proven track record of designing Python-based **data APIs**, building **scalable data pipelines**, and integrating **AI capabilities** into production systems.

WORK EXPERIENCE

Senior Data Engineer | Shopee

March 2023 ~ Present

- Managed large-scale **Airflow** workflows supporting both financial reporting systems and machine learning operations.
 - Built and maintained 100+ production pipelines processing 500GB+ of structured and model-related data.
 - Designed and implemented **CI/CD** workflows for automated DAG validation, **AI-assisted** code review, and deployment, ensuring version consistency and reducing manual release risks.
- Built and maintained internal backend systems serving 500+ users across finance workflows.
 - Architected RESTful APIs using **FastAPI** and Uvicorn, integrating Google OAuth-based SSO and implementing Role-Based Access Control.
 - Containerized services with **Docker** and deployed on Kubernetes, enabling auto-scaling, rolling updates and zero-downtime deployments for production stability.
 - Established centralized logging infrastructure using **ELK** stack and integrated Slack webhook alerting to enable real-time error monitoring and faster issue resolution.
- Built **LLM-driven** compliance agents to automate regulatory notice processing and product enforcement workflows.
 - Architected graph-based multi-agent flows using **LangGraph** to coordinate government document parsing, rule validation, and exception handling for product delisting and alert generation.
 - Processed 200+ daily government emails and 3,000+ products with 99.5% task accuracy in production.
- Streamlined business workflows by building internal platforms (FastAPI, Electron), reducing manual operational effort by 90% across 50+ users.

Machine Learning Engineer | Shopee

Sep. 2019 ~ Feb. 2023

- Designed and productionized a large-scale Similar Product Search platform serving 15M competitor products.
 - Developed a multi-modal embedding architecture (VGG + Sentence-BERT) with Faiss-based ANN indexing, enabling 10s-level large-batch similarity retrieval.
 - Engineered a GPU-aware distributed embedding pipeline with dynamic DAG generation in Airflow, implementing producer-consumer orchestration via **RabbitMQ** and async **MongoDB** storage to support scalable, high-throughput processing.
- Implemented an Image Copy Detection System based on ResNet, automatically scanning cross-border product images to identify and report copy issues, achieving 95% detection accuracy.
- Engineered an asynchronous, multi-worker distributed crawling system to scrape large-scale competitor product data, optimizing concurrency and I/O throughput to reduce overall crawling time by 90%.
- Designed a distributed task system (**Celery**, **RabbitMQ**, **MySQL**) to automate operational workflows, reducing a 2-month manual process to hours.
- Applied K-means clustering to optimize logistics endpoint planning, achieving ~80% coverage while reducing operational cost.
- Led model design for a push notification recommendation system (SVD), improving click-through rate by over 2x.

Data Science Intern | Shopee

Sep. 2018 ~ Sep. 2019

- Built automated data reporting pipelines, delivering daily user traffic insights to stakeholders and enabling data-driven decision-making.
- Performed data cleaning and exploratory analysis to identify key features, reducing model training cost by ~50%.
- Developed high-volume web crawlers in Python, collecting ~3M competitor records daily.
- Designed proxy pool and session emulation mechanisms, significantly improving crawler stability and ensuring consistent data availability.

SPECIAL EXPERIENCE

- Speaker (Talk & Tutorial) | 2022 PyCon APAC | 2022/09
- Lecturer | Happy Coding Online Course | 2020

EDUCATION

- B.S. Big Data Management | Soochow University | 2015 ~ 2019